BEING NICE ON THE INTERNET: DESIGNING FOR THE COEXISTENCE OF DIVERSE OPINIONS ONLINE

A Thesis
Presented to
The Academic Faculty

by

Catherine Grevet

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Interactive Computing

Georgia Institute of Technology
May 2016

BEING NICE ON THE INTERNET: DESIGNING FOR THE COEXISTENCE OF DIVERSE OPINIONS ONLINE

Approved by:

Dr. Eric Gilbert, Advisor
School of Interactive Computing
Georgia Institute of Technology

Dr. Amy Bruckman
School of Interactive Computing
Georgia Institute of Technology

Dr. Keith Edwards
School of Interactive Computing
Georgia Institute of Technology

Dr. Loren Terveen
Department of CS and Engineering
University of Minnesota

Dr. Ed Chi
Research Scientist
Google

Date Approved: April 1, 2016

To Grandpa

ACKNOWLEDGEMENTS

This dissertation is the fruit of many years of learning and questioning; countless times going back to the drawing board; and those few times of seeing the puzzle fit together. I'm so grateful to all the people who took time to answer my questions, to support me in moments of doubt, and to provide food for thought that pushed my research forward.

First and foremost, I am forever grateful to my advisor Eric Gilbert for having welcomed me in your lab later in the PhD process. Thank you for taking a chance on me. Aside from a brilliant mind I could always count on to answer critical questions about my work, your constant encouragement and positive attitude has been a driving force for me. Thank you so so much.

To my committee members. Amy Bruckman lent a hand out to me when I needed it the most; thank you for reaching out to me when I wasn't sure about my research direction, and that kept me going; you have provided incredible insights as I worked through my dissertation; you have made me a better researcher, and shown me how to be dedicated to ethics and integrity through your example. Keith Edwards was one of the first professors I met when I visited Georgia Tech; you have been a constant voice of kindness and availability throughout my time here; thank you for your support through the dissertation. Loren Terveen has been a formidable co-author and sound voice through my thesis work; thank you for believing in some of my risky ideas and encouraging me to push forward. Ed Chi helped me navigate through the process of publishing in industry, and showed me how a passion for research can help build large-scale systems; thank you for your guidance through my dissertation.

I am incredibly grateful for the constant source of help and support from great friends and research buddies among the years. To my labmates, Tanushree Mitra, Chaya Hiruncharoenvate, Saeideh Bakhshi, C.J. Hutto, I hope that our paths often meet again. One of the greatest resources we are told we have when we start the PhD program is the more senior students, Brian O'Neill thank you for your guidance. The PhD often feels like a lonely undertaking, but feels less lonely with friends who are there along the way. I'm especially grateful for the companionship and friendships of my roommates, Ying Xiao and Brendan Dolan-Gavitt, board game buddies Yee Chieh Chew and Liam MacDermed, conference buddy Elena Agapie, the 2009 GT CS PhD cohort, the IC PhD students for late night CHI and CSCW writing sessions (special mention to Takorea for feeding us), and to my Atlanta friends.

I would also like to recognize the people outside of Georgia Tech who have supported me and given me perspectives outside the PhD. To my Wellesley family, the amazing Computer Science professors and the alums I am lucky to still be able to count on today. To my colleagues during my Apple and Google internships, and especially to my managers Michael Coblenz and David Choi who were exceptional mentors. And finally, to my Yik Yak herd. My managers and coworkers have been incredibly supportive and helpful in giving me the time I need to balance both the PhD and my work at Yik Yak. I'm grateful for the flexibility and the kindness in recognizing this accomplishment, and special thanks to Derek Yang for having listened through my defense.

My family has always been there for me. Thanks to Mom and Pop for having helped edit this dissertation, it kept me on pace to finish and provided more encouragement than I can express. Thank you! My brother Jeremy for empathizing time and time again. And to my extended family for the centering I found in the wonderful holidays spent together, and to those we miss dearly.

Last, and most importantly, thank you so much Emile. For taking care of everything when I couldn't, for the times you explained my research better than I could have, and for always being the greatest cheerleader. You have given me strength and confidence, I love you and our little one.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Is social media better at tearing us apart than bringing us together? Despite connecting us globally, the Internet also brings with it the spread of uncivil or hurtful opinions. When conversations heat up online, people tune out or leave the conversation, and therefore do not end up hearing views that are different from their own, risking a society that is more and more divided. Users on social networks currently have limited options to remain engaged around highly charged topics, such as politics. Is this an inherent aspect of social media? Or are there design paradigms that could reverse this tendency?

In my work, I explore ways to design social media to facilitate more civil conversations, as well as two social computing research methodologies for building and evaluating social systems. As social computing researchers, we have a wide array of tools and methodologies at our disposal to understand how users experience social connections online. However, these are limited when it comes to asking questions about design alternatives such as: Would people use Facebook differently if they did not see any arguments in their Newsfeed? Since we cannot internally contribute code to modify the Newsfeed, are these questions out-of-reach for social computing research?

I met this challenge by developing a prototyping framework that I call *piggyback prototyping*, for social computing systems that require critical mass. With piggyback prototyping, I was able to leverage participants' existing social environments on Facebook, to explore novel interactions around civility on Facebook. I built a plugin for Chrome that modifies the content of a user's Facebook Newsfeed: it removes impolite posts, and highlights positive posts in green. This plugin was built on a politeness

classifier which was trained for Facebook posts.

To evaluate the plugin, I employed mixed methods in an *algorithmic probe* study, a variant of technology probes that actively engages participants in refining a social curation algorithm. Participants experienced the civility plugin on their Newsfeed for three weeks, and participated in daily diaries as well as a final interview. Through this study, I found that a civility plugin allows participants to perceive increased civility on Facebook. Highlighting polite posts was particularly compelling as a design direction in this space.

This thesis contributes to a better understanding of social media designs for more civil conversations online. I first demonstrated that people disengage from social media interactions when they encounter uncivil behavior from friends. To find alternative designs for social media that are more civil, I evaluated novel social interaction techniques. To do this, I designed a six-phase framework for prototyping social interactions called piggyback prototyping; and an algorithmic probe study methodology to include participants in the development of social curation algorithms. I built a piggyback prototype that modifies the civility on Facebook by highlighting positive posts in green and hiding impolite posts, and I deployed it as an algorithmic probe with 20 participants. I uncovered ways to improve the algorithm, and I found that participants responded most favorably to having civil posts highlighted. These findings open avenues for future research in designing pro-social platforms.

CHAPTER I

INTRODUCTION

My thesis focuses on building and evaluating systems to improve civility online.

Along with the rich potential of digital exchanges on social media comes the natural tendency to butt heads over viewpoints, resulting in many opportunities for hurtful, offensive, and uncivil remarks. One participant describes a common experience online:

> *Some of us got into a really big fight over [gun laws after Newtown]. And I defriended a couple of people. [...] So I tended to pull back from Facebook about that, but I mean not completely.*[1]

A long-standing and pervasive human behavior is to befriend and associate with people similar to ourselves. Alternately, we turn our backs, mock, shun, segregate, and, at the most extreme, eliminate those with whom we do not share the same affinities. For meaningful or superficial reasons, we are attracted to people who "get" us, and bicker with those who do not. Through social media, we have gained more opportunities to interact with people around us, as well as anywhere in the world, yet we have actually become worse at how we share ideas, opinions, and achieve consensus [19]. Consider the everyday scenario of browsing the morning news on your favorite media website, finding an article of interest, skimming through the comments section, and spotting a rude comment with which you strongly disagree. How do you react to this? Seeing opinions that conflict with our own is a common experience online, which the participants in my research describe as making them feel overwhelmed. In

[1] Taken from [40]

today's more globally connected world, our awareness of others has perhaps increased, yet it is still difficult to engage in a conversation across differences.

Is incivility inherent to online discussions? Is social media better at tearing us apart than bringing us together? Can we design platforms that allow us to carry on meaningful conversations despite differences? Can we build algorithms that present a broader variety of content? These questions have motivated my dissertation.

There is a large barrier to conducting this line of research: how can we build and test design alternatives to our existing social media platforms? If we want to test a new Facebook Newsfeed algorithm with the goal of obtaining a better pro-social outcome, we are faced with the issue that we cannot internally contribute code to modify the Newsfeed. Perhaps we could build a parallel research social media site, and at that point, have we not just recreated Facebook? Who would be interested in joining our platform? While much work in HCI, CSCW, Sociology, and Political Science have uncovered dynamics around polarization and incivility, exploring alternative design patterns for pro-social outcomes presents significant challenges. In my work, I contribute a solution to this problem.

I first set out to study the conditions under which diverse people can inhabit online spaces, and maintain relationships across their differences. Uncovering challenges around the notion of civility, I then set out to explore design alternatives for more civil conversations in social media. To do this, I conducted a prototype deployment intended to present conversations in a more positive light. A large portion of this work is the description of *piggyback prototyping*, a prototyping technique for social systems, which was used in the context of building a civility plugin for an existing social media platform. I then developed a study methodology called an *algorithmic probe*, which is inspired by technology probes, with the distinguishing goal of actively engaging users in the development and iteration of a social curation algorithm. Through this work, I found that encouraging positive behaviors for pro-social outcomes can make

participants more aware of the meaningful, uplifting, and motivational content shared by their friends.

1.1 The challenge of diverse opinions in online spaces

With a deepening gulf between political parties, and constant debates on many social issues, the United States is becoming more polarized and segregated by ideology [27, 6]. Polarization seeps into many other aspects of daily life: political parties promote world views that are orthogonal to each other [45]. For example, parenting choices are aligned with voting patterns [45]. Both online and offline, these political and lifestyle preferences steer community structures to create social echo chambers.

Segregating into like-minded groups can be harmful because we cannot hear the other side. Inwardly focused groups are unable to challenge their own views [73, 84]. This widening divide may cause even greater problems, such as the stalled process of choosing a new Justice of the Supreme Court due to tensions between Democrats and Republicans in the beginning of 2016 [50]. Encompassing a diversity of perspectives is valuable to any society, and to the principles of democracy. For example, exposure to diverse opinions engages society in healthy deliberation, which in turn creates a more informed society [53, 84].

Online we can connect to anyone in the world. Social media could be the platform by which we encounter different opinions, and engage in the democratic process of deliberation. In fact, online, we have more exposure to diversity than we do face-to-face [26, 67]. But this effect remains limited. Making connections with people similar to us has always been a part of human nature, and is a widely studied sociological phenomenon called *homophily* [60]. Our tendency towards homophily and effects such as the filter bubble, overpower the Internet's potential towards exposure to diversity. Indeed, social network analyses of blogs and Twitter have shown that we stay connected in groups of like-minded others [1, 33, 14]. There is untapped potential

here for online environments to go further to support the sharing of diverse views.

Cross-cutting conversations, or conversations that occur between two people of different opinion, do happen online [14]. Conover et al. found that Twitter replies between Democrats and Republicans are common [14]. However, a first look at these conversations reveals that they may be more irreverent than constructive. Noting this nuance points to a valuable consideration: simply increasing the volume of communication between two people of a different opinion will likely lead to misbehavior.

1.2 Approaches to creating more inclusive online spaces

Prior research in CSCW and HCI have explored ways to increase exposure to diverse views through novel social interactions (mechanisms that connect people through technological artifacts), and interface interactions (visual elements of an interactive artifact). In terms of social interactions, novel algorithms and interaction design can introduce people to others with different views [22]. Political Blend is an app that pairs two people of different political leanings and proposes they meet for coffee. Their work points to possibilities to explore further in introducing people across differences. The scope of the deployment for the app was inconclusive as to its impact long term, yet the fact that the system did introduce people across the political spectrum indicates the potential in this area.

In their work on presenting articles of different opinions, Munson et al. focus on interface elements that can provide a mix of political perspectives in news aggregators [64]. Prior work in this line of research found that some people were diversity seeking, while others were challenge averse, and these differences matter in the preferences of viewing news snippets. People who are diversity seekers like to see a balanced set of news snippets, while challenge averse users responded best when news articles that match their opinions were highlighted [64].

Using insights from this past work, I build upon it to understand how we can

design social media to foster more civil conversations. Incivility and the tone of online conversations are large obstacles to cross-cutting discourse. Social media can put two people of opposing views in the same room, but can it make them listen to each other? Design has the potential to deliver more diverse viewpoints, and designing for civility is a crucial step in increasing exposure to opposing views.

An unspoken reality in the implications for this line of study, is the challenge associated with obtaining critical mass. If agreeable information is more engaging and more appreciated by users, then systems that pair individuals with cross-cutting opinions might be unsustainable. In research, this could lead to projects stalling at the phase of enrolling enough participants in a study. From a business perspective, the financial benefits in catering to personal desires might come in conflict with the societal benefit of providing diverse perspectives. In this dissertation, I provide a framework for building social prototypes in order to leverage existing social media platforms, so that social systems can be evaluated in context.

1.3 Contributions of this thesis

This thesis makes contributions to Social Computing by increasing the understanding of design implications around designing for pro-social purposes; in this case social media designs that break sociological phenomena such as homophily. In addition, the contributions to HCI include the development of a framework, and examples for prototyping social interactions - an under-explored type of prototyping technique.

1. **A formative understanding of the implications of disagreements in social media.** Here, I studied the consequences of people having friends on social media with opinions different from theirs. I found that participants who had friends of different opinion from theirs engaged less on Facebook than those who perceived more homogeneity, during political events. Participants ended up unfriending or hiding friends of different opinions altogether, and described long

pointless conversations on Facebook that resulted in "agreeing to disagree". Finally, weak ties were particularly brittle to political disagreements, these friends were often viewed as irrational and were likely to be unfriended.

2. **A prototyping framework for evaluating large-scale social systems.** I developed a framework for prototyping social interactions in contexts that require critical mass called *piggyback prototyping*. Evaluating design paradigms around the notion of civility on Facebook requires such a prototype. I describe the steps of piggyback prototyping, provide two detailed accounts of using piggyback prototyping, and discuss the scope, benefits, and limitations of this technique.

3. **An algorithmic probe methodology for evaluating social algorithms.** I introduce *algorithmic probes* as a study methodology for actively involving users in the development of social curation algorithms. Participants were actively engaged with the probe and suggested ways to improve a politeness algorithm.

4. **An algorithmic probe evaluation for designing civil social media.** Using the piggyback prototyping framework and the formative findings, I deployed a plugin that made Facebook more civil for participants as an algorithmic probe. Participants who did perceive a change in civility were most influenced by seeing civil posts highlighted on a colored background, as opposed to another feature of the plugin that hid impolite posts.

This thesis opens up a number of different avenues for future research. It focuses primarily on friendship-based social media. The implications in the context of anonymous networks, or other network structures, merit their own research agenda. There are also many ways to expand on the framework for prototyping large-scale social systems. Finally, this work calls for further explorations around the notion

of emphasizing pro-social behaviors in social media rather than punishing deviant behavior.

1.3.1 Studying contentious relationships online

My dissertation aims to present an empirical understanding of current practices around engaging with friends of a different opinion. I studied how people currently manage relationships with those whose opinions differ from their own.

I looked at the intricacies of relationships between friends of opposing viewpoints in the context of political discussions on social media, because when they work out, they suggest the conditions under which diverse opinions can coexist online. People do hear opinions from the other side despite their strong tendency towards homophily [14]. One goal of this exploration was to uncover design implications for social media towards breaking down online echo chambers.

I looked at how Facebook usage and relationships were impacted by controversial political events, and found that weak ties were brittle, and that people tuned out of the conversations when there were too many opinions that were different. Even though people could simply terminate relationships with friends of dissimilar views on Facebook (and there were certainly those who did that), there were many connections that were able to be maintained. Through a combination of behaviors on Facebook like hiding, tuning out, logging off, or avoiding certain conversations, people negotiated around those differences to stay connected. The challenges faced by the participants suggest ways that social platforms like Facebook could better support these relationships.

Since I conducted this study, reports from Pew have confirmed these findings [62]. The growing research interest in polarization points to a pressing issue. From my research, I found evidence that incivility and a lack of respect were behaviors that resulted in people resorting to homophily-like behaviors. Through a prototype

deployment, I set out to uncover design implications for social media to facilitate civil conversations.

1.3.2 Prototyping large-scale social systems

Placing a prototype in people own social environment would shed light on the design implications for these systems. To understand the design implications for my project, I wanted to provide people with a prototype that they could use in their existing social environment. Other related work looking to present users with balanced news articles [64], or to counter more deviant online behaviors [20], have mainly been done in a lab setting. These studies have uncovered valuable insights, yet, the kinds of questions that can be explored in the lab, or a proof-of-concept deployment, are limited. I therefore developed a prototyping technique, called *piggyback prototyping*, that allows for an "in the wild" deployment within a user's pre-existing social environment.

Piggyback prototyping addresses the issues faced by social computing researchers when trying to evaluate a prototype. Many are faced with the possibility of needing to build their own social media platform. Some of the problems encountered with building an infrastructure from the ground up include: difficulties attracting users to obtain a reasonable critical mass, mitigating the novelty effect of a new system, and the potential for technical issues. To overcome these challenges, I developed a framework for social systems prototypes. This allows a researcher to focus on important research questions such as, why do users display certain behaviors, what motivates them, and what differences between users do we notice. Questions around civility and other pro-social behaviors online are especially ripe areas for utilizing this prototyping framework.

1.3.3 Interventions to support civility in social media

I focus on civility as a necessary component to maintaining cross-cutting friendships. The term civility encompasses a number of characteristics, such as conversation tone

and social norms [70]. Encouraging civility on Facebook could alleviate the frustrations I heard from participants when I studied the conditions under which they maintained contentious relationships during political events.

My work and related studies have shown the chilling effects of incivility on social participation [40, 54]. Could discouraging incivility in online discourse be the way to go? Limiting free speech on platforms that are inherently social may be too restrictive. Instead, can we create incentives that encourage civility?

I looked at a large aspect of civility, which is the tone of conversation, or politeness, which can be computed, and interactions can be built upon this computation. I built a browser plugin based on a classifier of politeness, and deployed it as a technology probe, engaging social media users in a broader discussion about social norms around civility online. The plugins are not meant to be perfect deployable products.

The main building block for these plugins is a politeness/impoliteness classifier. I built it based on existing computational linguistics work [16]. This classifier considers 20 features (such as the presence of "please" or "sorry"), and categorizes a given sentence as polite, impolite, or neutral [16]. I trained this classifier on Facebook posts, and released it as a browser plugin to participants in a technology probe study.

Through a mixed methods approach of surveys, interviews and diary studies, I found that encouraging pro-social behaviors, such as highlighting civil posts in green, was appreciated. The participants in my study who did perceive a change in civility on their Facebook Newsfeed liked seeing civil posts displayed more prominently. There is no indication that hiding impolite posts has an effect on perceived change in civility.

1.4 Thesis overview

This thesis is composed of three main parts, and each focuses on a different study related to its contributions:

- In Chapter 2, I present related work in the areas of Sociology and Political

Science, as it relates to homophily, polarization, and the consequences of these behaviors on society. I also describe systems that have been built with the goal of achieving pro-social outcomes.

- In Chapter 3, I describe a study on friendships between people of different opinions during political events on Facebook. The findings from this formative work guide the rest of the dissertation.

- In Chapter 4, I describe the piggyback prototyping technique for large-scale social systems I developed, in order to be able to build a prototype for the results from the Facebook study. This section presents the technique, and I provide examples of uses for it.

- In Chapter 5, I present the results from the culminating study of this work, where I conducted a longitudinal diary and interview process with participants, to determine design guidelines for building social media in a way that encourages civility.

- I conclude with future research directions paved by this study, and a discussion of the limitations of this type of research.

CHAPTER II

LITERATURE REVIEW

From an empirical perspective, my research has drawn on the work of political scientists and sociologists exploring polarization, and exposure to diversity, among the general public. Along this theme, social computing researchers have specifically studied how that occurs in social media. In my formative work, I found some evidence of a link between the prevalence of incivility in social media, and the growing trend of polarization. Applying social computing research methods to these areas can provide insights into ways in which civil social media could be designed to allow people to hear from a diversity of perspectives.

Second, my work requires functional artifacts in order to understand the implications of novel system designs. As such, systems building and prototyping are integral aspects of this type of research. HCI has a long history of developing prototyping methods, yet, social computing, which is arguably more difficult to prototype, does not have much guidance in terms of building adequate prototypes. Thus I will also lay out different prototyping tools and areas in HCI as they relate to my research, and provide some background for the contribution of my work in terms of social computing prototypes.

In this Chapter I will present related work demonstrating the issue of polarization and the sociological behaviors that influence it, and then I will elaborate on civility and discourse in particular, as it plays out in social media. Finally, I will review literature in prototyping and systems building for pro-social applications.

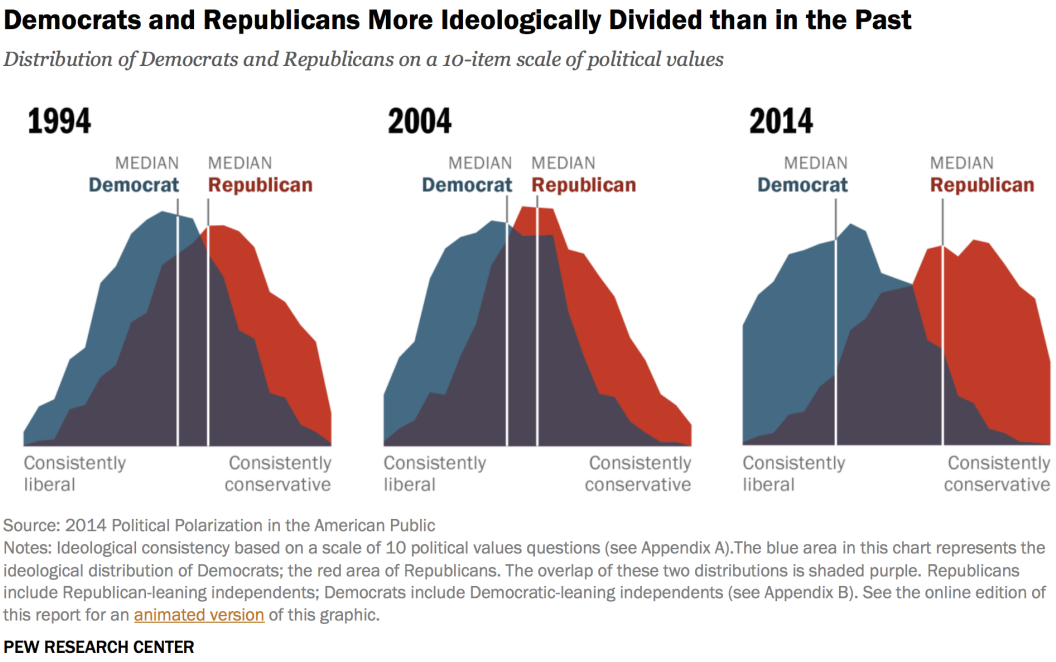

Figure 1: Growing polarization between Democrats and Republicans [19].

2.1 Polarization and homophily

Polarization in politics is growing in the United States [27, 6, 19]. In a Pew Research Center study of 10,013 American adults, researchers found that the percentage of those gravitating to core partisan values has doubled in the past two decades (from 10% to 21%). The result is a growing gap between the parties (see Figure 1).

Adhesion to more and more extreme political party views is accompanied by growing negative views towards those in the opposing party. In 1994, 16% of Democrats had very unfavorable views towards Republicans, and 17% of Republicans had very negative views of Democrats. Twenty years later, in 2014, they found those numbers had risen to 38% of Democrats, and 43% of Republicans [19]. It is unclear what is the cause and what is the effect: does an increasing gap in political agendas create a deepening gap between people of opposing political views? Or is it the increasing animosity that leads people to react with more extreme positions? Nonetheless, we see this vicious circle causing deep misunderstandings throughout the country.

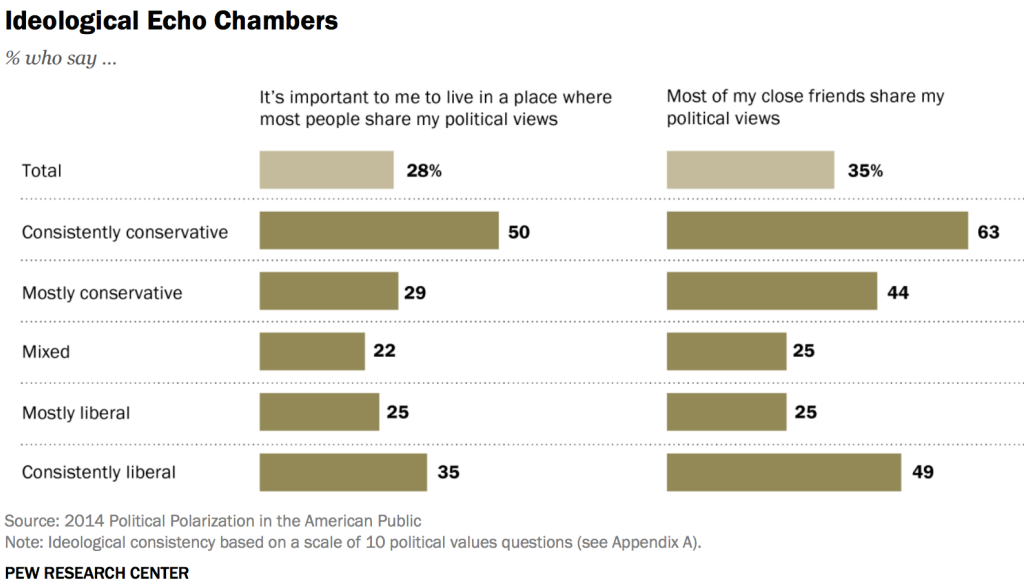

Figure 2: Attitudes towards homophily based on political preferences [19].

The issue of polarization goes beyond politics, as lifestyle choices and personal values align with political preferences through a process called "the big sort" [45, 6, 19]. Bill Bishop describes the "big sort" as the process by which people choose to live in certain cities or neighborhoods because of desired lifestyles, which separate groups along political lines [6]. Homophily, and certain reinforcing behaviors such as social influence, by which certain prominent individuals in a community shape the thinking of their followers, and selective exposure by which people choose information that reinforces their views [30, 49], lead groups to be homogenous and lack diversity of experiences and of thought. There are certainly benefits to congregating with like-minded others, such as strong emotional support [60], and a better ability to mobilize partisans [60]. On the downside, this tendency can magnify already extreme views, and increase the divide between groups of different opinions resulting in "ideological silos" (see Figure 2) [64, 83].

While polarization may be blatantly evident in data about political parties and the public alignment with those ideals, my interest in designing social media to bridge across differences goes beyond politics proper. For example, the choice of living

conditions is not the same for people in the different parties, and they self-segregate into neighborhoods of the same political leaning [6, 19]. The study showed that Democrats tend to prefer living in small houses close to each other and in cities, while Republicans lean more towards preferring larger houses further apart and living in rural areas [19]. Political discussions are certainly a prime area to study homophily in social media, however these views seep into other daily conversations. Designs that enable people from different political perspectives to be exposed to other ideas could go beyond politics, and inform designs to bring people together across a broad array of differences.

2.1.1 Social capital and bridging capital

Polarization can be harmful if it prevents us from cohabiting or collaborating with people of a different opinion from ours [73, 84]. Communities are better able to fight crime, share resources, and achieve a greater quality of life when they are enriched by strong *social capital* [73]. Social capital refers to the intangible transactional aspect of personal relationships. Building social capital requires trust and an ability to depend on others. A society divided by polarization may lack the necessary social capital amongst its citizens.

Maintaining social capital in a divided society requires individuals willing to cross ideological boundaries. In social networks, these individuals are *network bridges* that cross *structural holes* [13]. They are in a unique position to convey information from one group to another [13]. Structural holes are characteristic of social networks, and are present regardless of other social pressures that might divide social groups. Understanding the complexities of this network position can offer opportunities to facilitate conversations across different ideologies.

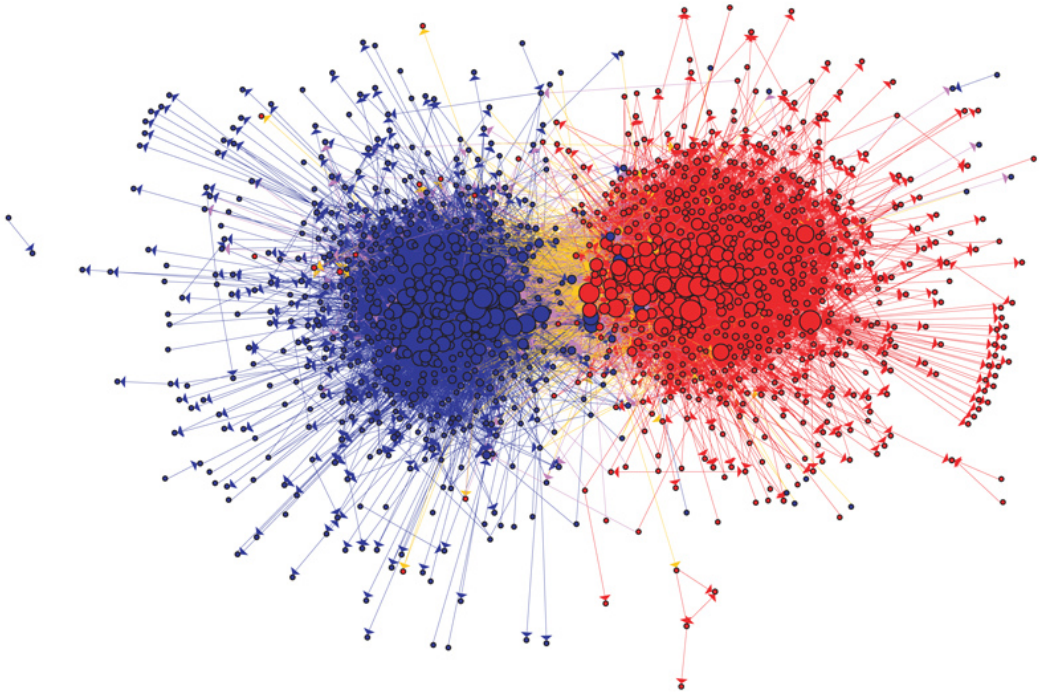

Figure 3: Adamic and Glance found that blog authors linked only to others of a similar political leaning [1].

2.1.2 Exposure to different opinions in social media

While the Internet and television are better at exposing people to differing opinions than real-life encounters [67, 26], the facility with which one can construct a comfortable online environment through homophily and selective exposure may contribute to the growing divide. In fact, the online experience particularly facilitates the process of homophily: recommender systems and the algorithms employed by many social sites present users with people and posts that are agreeable to their preferences [22]. Systems in widespread production often reinforce our tendency to associate with others like us (e.g., Amazon's product recommendations, Facebook friend suggestions). Similarly, mechanisms amenable to recreating offline social networks online, also provide the ability to further "weed out" those of a different opinion, such as through blocking, hiding, or unfriending.

As seen in Figure 3, the links in blue (Democrat), and red (Republican), are more

strongly connected than the sparse yellow and purple links in between. Work by Gilbert et al. confirms this by showing that there is an "echo chamber" effect in blogs [33]. Studies of Twitter have shown that, similar to more traditional blogs, people do not link to opposing political opinions through "retweets", a function of Twitter that lets users spread a message to their followers [14].

While television is still the place where most Americans get their news about politics (49%), Facebook is just behind (48%) [62]. These results were found by a Pew survey of people who use the Internet. If these numbers were at the scale of the wider population, this would be 39% of all Americans, which is still a significant fraction of American adults [62]. People get a large portion of their political news online. Aside from Facebook, 14% reported getting political news in the past week from YouTube, and 9% from Twitter [62]. The role of social media in providing information and mediating those "healthy debates for democracy" is consequential.

Social media and the Internet does put people in the presence of dissenting views [62, 25]. On Facebook, users are exposed to an average of about 20% of news articles that are cross-cutting from their self-reported political affiliation [25]. However, this might depend on the intensity of people's political engagement, and their ideological perspectives. For example, those who are consistently conservative tend to have a single news source (47% of them citing Fox News as their main source of news), are more likely to have friends similar to them, and have more friends on Facebook who agree with them. On the other side of the spectrum, liberals consistently access a wider variety of news sources, but have a higher tendency of unfriending on social media, or ending a personal relationship because of politics [62].

2.1.3 Disagreements in social media

While our online environment is strongly homogenous and agreeable to our preferences, online social networks are not entirely homogeneous [25, 1]. Disagreeing with

a friend online is not uncommon: 73% of social media users report disagreeing with a friend's post [76]. In fact, people tend to overestimate similarities with their friends [35]. Furthermore, half (47%) of conservatives, and (59%) of liberals say that they sometimes disagree with their closest political discussion partner [62]. For those who have more mixed political views, the number rises to 79% [62]. About 20% of political content shared on Facebook comes from friends of a different perspective [25].

Social media users are exposed to opinions that differ from their own more accidentally than purposefully, meaning that they see a post containing an opposing opinion through a friend posting it, rather than specifically seeking it out [65, 76, 87]. This observation is characteristic of people who bridge across structural holes in their social graph, meaning that they connect two social groups that don't have many other connections between them [13]. For example, about 25% of postings on non-political blogs are of a political nature [65]. This means that a blogger who usually posts about travel, may once in a while post about politics and thus exposing her followers, who may be primarily interested in travel, to some political thought.

When political tensions arise in these non-political places, people tend to employ ad-hoc mechanisms to minimize animosity [33, 65, 23]. For example, political posts on non-political blogs often contain warnings such as "please excuse my rant" [65]. Another relevant mechanism is self-censorship, or refraining from posting [17, 79, 88]. More drastically, realizing that a friend holds a different opinion has led 18% of social media users to unfriend, block or hide someone [76]. Indeed, the larger culprit for the lack of exposure to dissenting perspectives is more attributed to individual behaviors, than to the algorithmic curation of the Facebook Newsfeed [25].

Cross-cutting conversations may exhibit different patterns, depending on whether they occur between strong ties or weak ties. Tie strength is the concept that different relationships have different levels of intimacy [37]. Close ties are people who are often family members and close friends, while weaker ties are acquaintances from various

social circles [37]. Our social environments are composed of a variety of tie strengths, and they confer distinct benefits: strong ties provide more emotional support, while weak ties are important for exposing us to new information or help during career transitions [10, 37]. Tie strength is an attribute of our relationships in person, as well as the ones we maintain online [34, 10].

Another reason cross-cutting conversations may differ depends on the privacy context of the platform. Most research available on network structures of political conversation draw on public discussions, since that data is more readily available. Conover et al. found that cross-cutting conversations happen privately, perhaps due to what people are willing to display publicly [14]. They showed that the "mention network" (a network of semi-private messages) on Twitter between Democrats and Republicans, is strongly connected [14]. This means that the communication that is happening across parties is happening privately, rather than being publicly displayed. However, these conversations may be more irreverent than constructive, though a sentiment analysis remains to be conducted on this dataset.

2.1.4 This work

In this work, I study relationships between friends of different political opinions, because examining their success suggests ways in which we could better support them. I conducted an exploratory study around this notion, which I describe in Chapter 3. I found that one big problem which happens during times of political tension, is friends tuning out of conversations. The feeling of being overwhelmed by discourteous behavior during political events pushes users away. While this behavior may be the driving force behind increased polarization (rather than algorithmic filtering), it is still the tools and affordances provided by social networks that set norms and facilitate behaviors. From this, I looked deeper into the notion of civility and ways to build more civil social media.

2.2 Civility and politeness

The central building blocks of my thesis are founded on the notion of civility as an important component to society. The term *civil* is quite vague and subjective, in contrast to politeness, which is more straightforward and more easily computable. In this work, I use both terms in tandem: *civility* as the ultimate goal encompassing all attributes necessary to collaborate and cohabitate as a society; *politeness* as a more concrete and actionable attribute of civility. Further explanation will clarify the difference between the two.

2.2.1 Defining civility and politeness

Defining civility is not straightforward [70]. The term "civility" comes from the latin "civilis" which is also the root of "citizen". As it pertains to its origin, the term "civility" refers to the concept of being able to live together in society. Framing it in the context of Goffman's "theory of face" can help to understand civility in social media discourse. The theory of face describes how people engaged in a conversation strive to maintain their social adequacy, and help each other through certain cues to avoid humiliation or embarrassment, thus preserving dignity [36]. Civility can be characterized as this desire to maintain face. By following the norms of civility in society, an individual can maintain a proper face amongst other citizens. While Goffman's theory was constructed prior to the Internet, it is key to understanding the disconnect between people's desires to maintain face offline, and the inherent difficulty of achieving this online. Indeed, understanding proper norms in an online environment, which is opaque to social cues, might be more challenging than face-to-face.

Civility may commonly be synonymous with politeness, though the two terms should be disentangled [70]. While politeness is a component of civility, the term civility refers to a broader notion of democratic morality [70]. By this I mean that

a civil conversation is guided by democratic values in which a collective wellbeing is upheld. In contrast, politeness refers to a set of accepted stylistic communication rules, or the means of communication [9, 16]. Politeness can aid civility. Yet, incivility can also occur without impoliteness [70]. In this dissertation, I distinguish between civility and politeness, in that I refer to civility as an overarching goal of this research on a societal level, while politeness is the computable aspect that I employ through my prototype design.

There are some positive effects of incivility, such as increasing engagement. Uncivil comments on blogs encourage people to participate more (as compared to civil comments) [8]. On television, uncivil debates draw a large audience [66]. It is a fact that "bad" behavior is entertaining [71]. In a world in which view count and engagement metrics are the ultimate drivers behind the design of social media, toning down this entertainment might not make business sense. There may be a challenge in making the tone of online discourse more civil, and keeping it engaging.

Flaming, or personally attacking an individual in an online forum, is an example of an online behavior that tends to be thought of as uncivil [70]. Social media platforms could detect patterns of behavior akin to flame wars or other deviant behavior. However, taking a blanket approach to these phenomena might be unwise. Indeed, these types of behaviors are not necessarily uncivil. For example: would calling out someone making racist remarks be considered uncivil? No, as it supports social justice and societal well-being. In fact, a pro-social system should help make those voices heard.

Incivility devoid of impoliteness is the most threatening, according to [70], because it may concern a deeper form of disrespect. Incivility has direct consequences for cross-cutting relationships. It has been shown that in televised political debates, incivility increases negative feelings towards the other side [68]. In September 2013, Popular Science removed their social features - namely the comments sections - on

all of their articles, because of the negative effects of incivility on the user experience around their articles [54]. While disagreement is necessary in a healthy democracy, alienating arguments result in the current culture wars [63].

2.2.2 The role of conflict in discourse

Providing more opportunities for civility does not preclude that conflict and anger are important components of discourse, especially as they pertain to issues around civil rights. When pacifists choose voice over violence, asking them to always remain civil through this discourse is an unrealistic expectation. Martin Luther King, Jr. expressed the duality between tension and justice in his letter to the Alabama clergymen: "Nonviolent direct action seeks to create such a crisis and foster such a tension that a community which has constantly refused to negotiate is forced to confront the issue" [51].

Does upholding civility stifle minority voices? One could argue that, while a civil rights movement can contain impolite language, it is by definition a civil discourse because it upholds civil rights, equality, and democratic principles as the ultimate goal against biased elites. The latter are the actual ones being uncivil, despite the fact that they have the privilege of polite conversation.

Asking for civility in conversation does not mean that there will be no conflict, which in fact is necessary to civility. Instead, conflict is more easily dealt with in an environment that supports civility, since ideas can transition between people of different ideals. In a world without civility, people are much more likely to conform to their own social circles - much like what we see currently happening in the United States.

2.2.3 Social norms in social media

When taken to its extreme, uncivil behavior online is called cyberbullying, and is a widespread issue today, especially among teenagers. Cyberbullying can happen in

many different ways, such as targeting, doxing, and others [58]. Bullies may employ sophisticated schemes on multiple platforms, and apply considerable resources to harm a specific target. Cyberbullying can lead to grave consequences for the victim, such as psychological trauma, self-harm and even suicide. Its display and ramifications can differ almost on a case-by-case basis, with its worst expression in the example of criminals exhibiting psychopathic behaviors to purposefully target a specific person.

Because it is a widespread issue in social media, social computing researchers have looked into cyberbullying, and created tools to assist the victims and friends of victims [58, 20]. For example, Dinakar et al. propose a prototype for a system that gives a time out before posting in a heated argument [20]. The prototype creates a delay before something can be posted, in order to give the poster a moment of reflection, and slows down the pace of a debate. Studies aiming to address cyberbullying in social media can provide insight into what might work in a less intense situation of uncivil conversation. For example, while time out seems to be a great idea for cyberbullying, it could also be a valuable method in handling disagreements or heated conversation, much like the common wisdom of waiting a while before sending an angry or emotional email.

There are certainly differences between incivility in social media, and cyberbullying in social media. Cyberbullying is, perhaps, more punctual and much deeper in terms of the resources deployed to conduct it (which also goes for the resources required to dismantle it). In the case of incivility, the person being uncivil may not even be aware how they are being perceived. They might not have the intention to harm, thus a punishment for their behavior might be exaggerated.

Less extreme, but perhaps more pervasive, are instances of disrespectful language usage. While freedom of speech protects legal action against what someone says, freedom of speech does not mean speech without consequences. There are repercussions for being rude. In moderated online communities, these norms are shared through

FAQ's (such as on Usenet) [80], or a Code of Conduct.

Reprimands can range from private emails, to public censure, or hellbanning [80]. Proponents of free speech might argue that removing incivility from discourse is a breach against the United States Constitution. Upholding civility as an optimal outcome might make some people think that what they want to actually express is being censored.

On platforms such as Facebook, which are not moderated by a common FAQ, this is more difficult. Machine learning tools can detect when incivility or disagreements arise in social media, and mediate those conversations properly [33]. This type of language detection can also help detect the credibility of sources. Through providing feedback about the tone in a discussion, people may alter their communication style.

2.2.4 Computing civility and politeness

Since politeness is a large component of civility, and more computable at this point, I will develop social media tools that engage participants around politeness online. From this starting point, I will then explore the experience around incivility. Thus, the goal of my work is to determine the best ways to design for civility given these nuances.

Incivility has direct consequences for relationships with others of different opinions. It has been shown to have a direct impact on the polarization of audiences of televised debates [68]. The Internet is a particularly uncivil place. For example, politicians on Twitter use alienating language more often than positive behavior, such as thanking people [44]. The experience around incivility online is an under-explored area in HCI and CSCW. While some studies have looked at ways to operationalize certain aspects of incivility, such as politeness/impoliteness [9], there have been no interventions in this area.

Danescu Niculescu Mizil et al. created a politeness classifier for content in social

media (on Wikipedia and Stack Exchange in particular) [16]. This classifier contains twenty features that account for sentence structure and content, in order to determine a politeness score (polite, neutral or impolite) for a given request. This work demonstrates the ability to compute politeness in a social media context. My work extends these findings by building usable systems that leverage this classifier.

2.2.5 This work

Throughout my dissertation, I refer to the notions of civility and politeness, where civility is the ultimate goal, and politeness is an actionable component of civility. I use a politeness classifier to parameterize online text as polite or impolite by adapting work from [16]. The system I developed was evaluated as a deployed "in the wild" technology probe to engage with the participants about civility online, going beyond the single notion of politeness.

2.3 Building pro-social platforms

In this section we look at work in the area of HCI and CSCW, and some theoretical underpinnings of these areas to explore how people have gone about creating and evaluating platforms that try to achieve pro-social benefits.

2.3.1 Prototyping social systems

In HCI, a prototype is "a concrete representation of part or all of an interactive system" [3]. This is in contrast to an abstract representation, such as a verbal description. As such, prototypes can be manipulated. Their manipulation creates a vehicle for communication between designers, engineers, and users [3]. The prototype should serve a well-defined function, such as prototyping the role of a new capability, the look and feel of an existing concept, or the implementation of how a system actually works.

Through the process of adapting and refining the prototype, it evolves. To allow for this flexibility, prototypes tend to be rough and sketch-like. Low-fidelity prototypes are a rapid prototyping technique: they are easy and quick [3]. Common methods are paper prototyping, where an interface is literally drawn on paper [78], or wizard-of-oz prototyping where the interface is a curtain behind which a researcher responds to user input [3]. Higher fidelity prototypes require significant coding, but provide a more finished look [3]. Crafting a relevant prototype requires the skill of a domain expert.

Prototypes serve to elicit guidelines for future design improvements, and they are evaluated in observational settings. To use Computer Science terminology, they are compared against "benchmarks of performance". These metrics are determined prior to a study, and guide the tasks that users will be asked to perform [3]. Other forms of evaluation are less directed: the user creates meaning through interacting with the prototype, such as with probes [31]. In addition, prototypes can also be viewed as essential tools for generating design ideas and insights [59].

Here we use "social computing systems" as a broad term for technologies that involve human-human interaction. We distinguish between two types of social computing systems: those for small groups, and those for larger crowds. Small-scale collaborative group systems can be prototyped with adaptations of traditional HCI techniques. Social computing research has many examples of these [38]. For example, "paratypes" are probes that can help understand the social context and social acceptance for a new technology [47]. When using a paratype, a researcher surveys reactions to the prototype as people go about their day-to-day activities. The study of the prototype is done in a way that is situated within the offline social and environmental context of the artifact [47]. To our knowledge, however, there are no prototyping techniques for social interactions online.

A simple search in the ACM digital library underlines this gap: only 4 papers

from the CSCW conference contain the words "prototype" or "prototyping" in their title, as compared to 85 papers for CHI. If we naively align HCI work to CHI, and social computing work to CSCW, it becomes apparent that prototyping techniques for social computing are scarce. This may be due to a number of challenges.

One major obstacle to building social computing systems is the need to obtain critical mass [41]. To avoid this issue, some systems have foregone testing with people, and opted instead for agent simulations [29], or simply proposed a prototype without evaluation [59]. In these approaches, it is not possible to evaluate the social affordances of the system. Others have created complete systems in the hope of attracting users to a polished product. Turkopticon, a tool for Amazon Mechanical Turk (AMT) workers, for example, used bootstrapping by collaborating with an AMT employer to obtain their initial users [48]. At the time of publication, four years after the launch, they had 7,000 installations [48]. Would their results have been irrelevant if they only had ten users? Some may be lucky, such as Link Different, a social translucence tool for Twitter, which gained 144,232 users in two months thanks in part to media attention [32]. In both cases, obtaining critical mass was central to their efforts and ultimate findings. Other projects may not be as lucky, or as able to devote efforts to obtain users.

Another challenge is determining evaluation metrics. Is a system that gets fewer than thousands of users a failure? Number of users is a common metric for commercial products, and we have also, perhaps to our detriment, adopted it as a standard for research. If achieving critical mass is our main measure of success, we are limiting the quantity and quality of social questions we can ask.

2.3.2 Prototyping social algorithms

Social media sites employ algorithms to curate the volume of information accessible to us [24, 75, 55]. Often, these algorithms are black boxes that are based on interactions

and inferred preferences of the user. Users of these systems are not generally aware that there is an algorithm curating what they see on Facebook [24]. Even less so do they have any insights into how their behaviors, what they post, and who they befriend affect the algorithm.

Should algorithms such as one that filters civility live in the algorithm of a platform like Facebook? Or is it more amenable as an external tool to put in people's hands, allowing them to choose to use it or not? These are questions that have not been explored, but pose important questions in relation to building systems that modify social interactions. In particular, as it relates to pro-social behaviors like civility, how do people react to being punished by an algorithm? How do we create a sense of trust between the user and the algorithm? How important is it to ensure that users feel they have control over these systems?

2.3.3 Systems that dismantle polarization

Work in CSCW and HCI looks into bridging across these political differences [22, 35, 53, 64, 82]. Novel algorithms and interaction design can present a mix of political perspectives in news aggregators [64], or can introduce people to others of different views [22]. Some of these interactive systems have demonstrated the potential to support political deliberation [53, 82]. For example, systems can support weighing multiple sides of issues [53]. However, systems that modulate the tone of conversation in social media have not yet been explored.

Building systems that break homophily are not by definition systems that aim towards peace and consensus. Adversarial design argues that agonism, or conflict between ideologies, is necessary for a healthy democracy [21]. Through this lens, design can be confrontational rather than supporting agreeableness [21]. It can support discussions, arguments and critiques, which all contribute in a healthy way to democracy. Yet, this conflict shouldn't be alienating to the point of creating segregated

communities that can no longer find common ground. Hearing the other side can have a damaging impact on relationships [67] and political engagement. Exposure to cross-cutting opinions can make people more ambivalent about politics [67], and may limit people's willingness to take political action [26], meaning that people who are more exposed to friends and family of differing political opinion may be less politically engaged than those who stay within their in-group. These considerations need to be taken into account.

For a person's self-development, knowing and interacting with those of different perspectives is critical [84]. It is how one grows, matures, and experiences the world around them. As a community, hearing the voices and perspectives of its constituents is vital. It is what democracy stands for [84, 6]. Furthermore, confronting diverse views brings about new sources of information, and leads to more educated decisions [53, 84]. Consequently, a community thrives when these different perspectives can be discussed and deliberated in a public forum. The point is not to require everyone to leave with the same viewpoint, but rather when people become aware of new ideas, they can better understand their own. For example, as the United States is a new home for many immigrants, being able to work together as a diverse society requires an understanding of those around us, their needs, desires, and beliefs [74].

Certainly, in a society that encompasses a diversity of perspectives, disagreements will inevitably arise. In fact, the theoretical framework of Adversarial Design supposes that without disagreements there is no democracy [21]. As a necessary condition, disagreements should be incorporated into design, rather than shut-off by design. Decreasing polarization should not come at the expense of stifling debate.

2.3.4 This work

With all of this in mind, I developed a framework for prototyping social computing systems in order to build a prototype for a social media platform that is more civil.

This framework can be used in a variety of different contexts, with the underlying goal of evaluating novel social interactions, as opposed to traditional interface interactions.

2.4 Thesis research questions

R1: Under what conditions do different opinions coexist in social media?

This question is addressed in Chapter 3, looking at Facebook users and how they manage relationships with friends of different opinions from theirs during contentious political moments. We found that incivility in social media results in people tuning out of the conversation.

R2: What methods can we use to build better social platforms?

In Chapter 4, I describe a prototype framework to build and evaluate social systems.

R3: What design components create a more civil experience on social media?

Finally, in Chapter 5, and based on the findings from Chapter 3 and Chapter 4, I study the impact of a more civil platform on people's experience in Facebook, from a longitudinal perspective.

CHAPTER III

STUDYING CONTENTIOUS RELATIONSHIPS ONLINE

In the first quarter of 2013, I conducted an investigation of heated political discussions via social media [40] [1]. The intricacies of relationships in the context of political discussions on social media suggest the conditions under which diverse opinions can coexist online. Three contentious federal events rattled the U.S. political landscape in 2013: budget cuts, gay marriage debates, and gun control regulations. In this political climate, I studied how Facebook users managed relationships with people who held different opinions. This study employed mixed methods through a survey and interviews. I obtained 103 survey responses, and 13 phone interviews about Facebook usage and friendships during these events, amongst Facebook users who have strong political opinions. This study revealed four insights:

- I found that participants who perceived more differences with their friends engaged less on Facebook, than those who perceived more homogeneity.
- Second, choosing *when* or *when not* to hear from a friend was difficult to manage, and participants ended up unfriending or hiding friends of different opinions altogether.
- Third, participants commented that they would get drawn into long pointless conversations on Facebook that resulted in "agreeing to disagree".
- Finally, weak ties were particularly brittle to political disagreements, despite being the ties most likely to offer diversity. These friends were often viewed as

[1] This work was published as C. Grevet, L. Terveen, and E. Gilbert. 2014. Managing political differences in social media. CSCW.

irrational, and were likely to be unfriended.

3.1 Research questions

In my Facebook study, I wanted to explore the conditions under which cross-cutting relationships could be maintained in social media. In particular, the overarching research questions were:

1) How do perceptions of differences affect engagement on Facebook?
2) How do people manage relationships with friends of different opinions?

These questions were studied in a specific and ever-evolving political context, as well as a social media context. Our survey and interview questions were dependent on the specifics of these systems at the time of our study, in order to elicit grounded responses from our participants. In the discussion section, we generalized from these contexts to highlight what can be learned for future designs of social media for political discourse.

3.2 Site of study: Why Facebook?

On Facebook, users can post status updates, pictures, videos and articles, and comment on friends' posts (see Figure 4). The main layout for the site is a Newsfeed that displays the most recent and noteworthy posts within one's network. An unknown algorithm sorts the posts in the Newsfeed, possibly prioritized by posts of high potential interest, and posts from close ties. Personal pages can be customized with a picture and information such as birthday, religion, political affiliation, and interests. Facebook includes a number of privacy controls to allow users to specify audiences for posts and pictures.

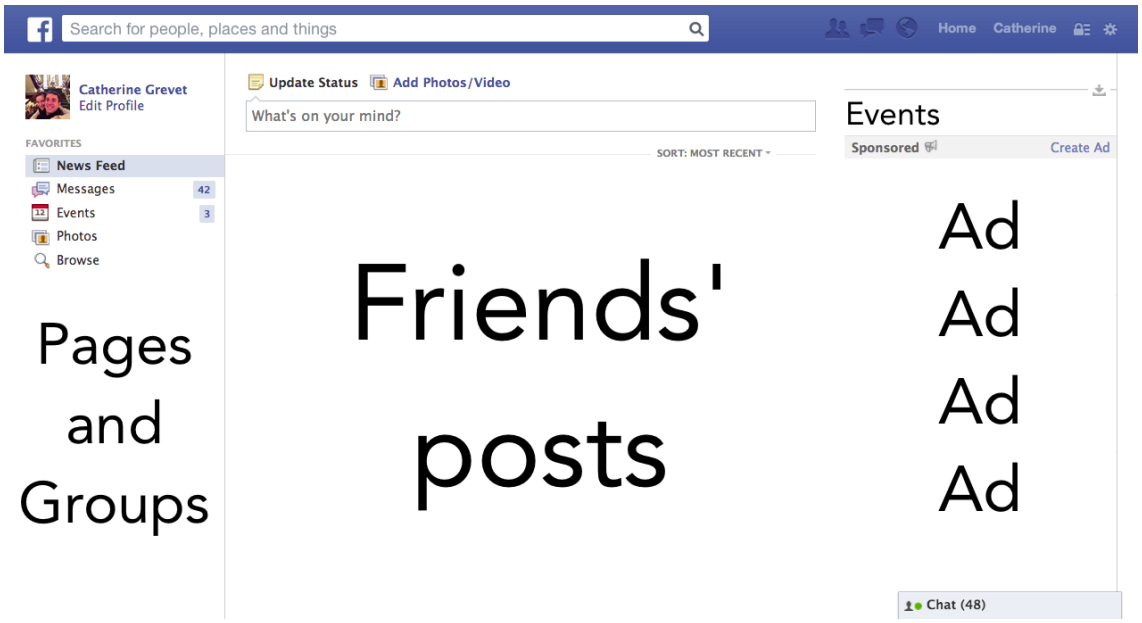

Figure 4: Facebook interface.

Facebook is the largest social network with currently 1 billion users [2]. This means that findings pertaining to Facebook have a large impact, especially as a site of choice for this study on friendships and politics: people get more political news on Facebook than from any other social media site (see Figure 5) [77].

Connections on Facebook are reciprocal and mostly mimic offline networks, as people are primarily connected to people they also know offline [7, 11, 25]. The impact of Facebook on different types of relationships has been widely studied [10, 34, 12]. Relationships on Facebook are expressed and maintained in ways similar to how they occur offline [34]. This has enabled researchers to study longstanding sociological theories through Facebook data, such as the fact that weak ties are helpful during job transitions [10].

What you talk about on Facebook has a powerful effect on your network [52]. In a 2012 study, Kramer found that users on Facebook can affect their friends for up to three days after posting a positive post [52]. They found that a friend would

[2] http://newsroom.fb.com/company-info/

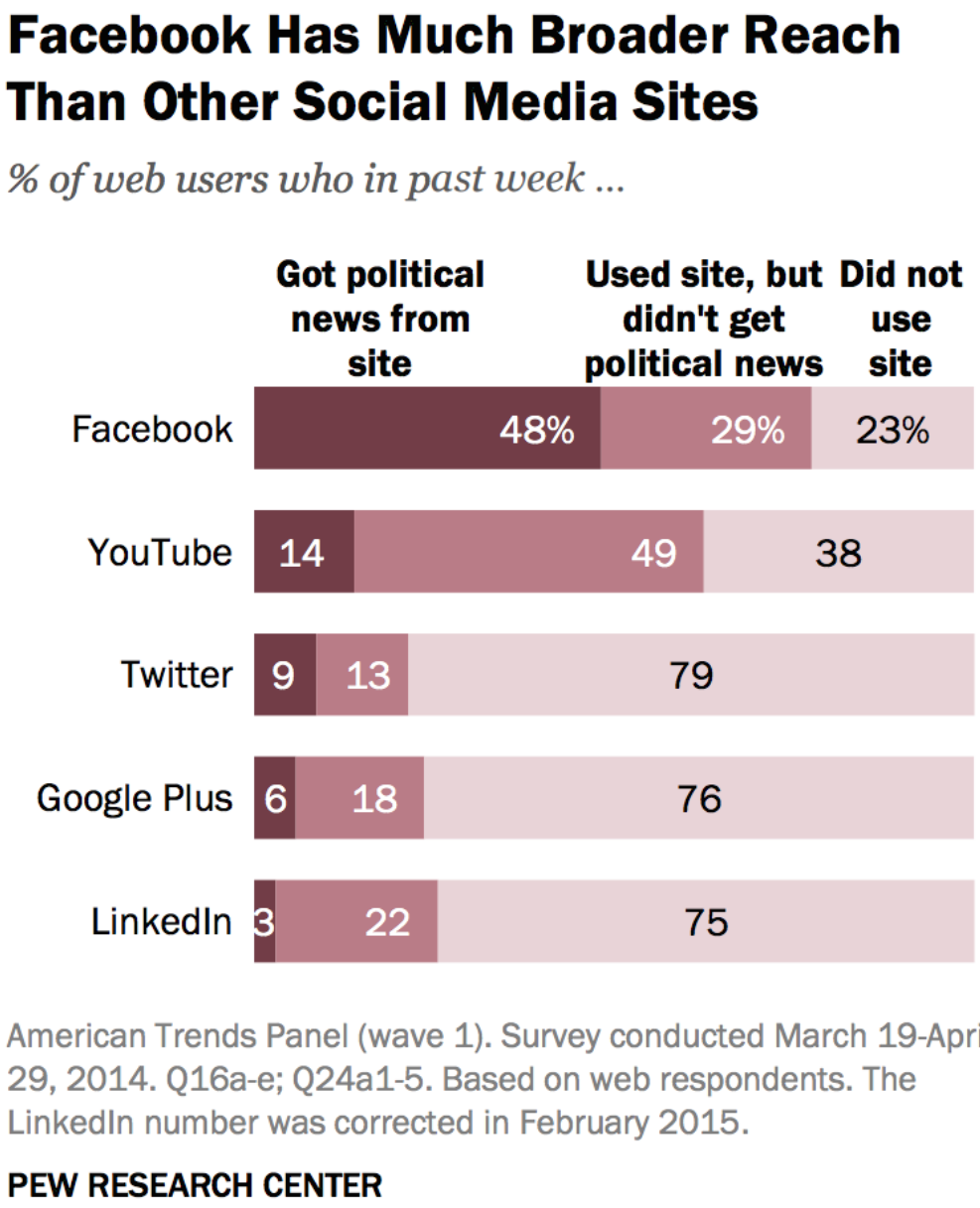

Figure 5: Usage of Facebook for politics compared to other websites from [62].

be more likely to post positive content after seeing a friend post something positive, whether this actually affected the mood of the friend or is a matter of social pressure is unclear. At the scale of Facebook, effects of this nature have a large impact, and make us wonder about the implications of negative comments in the context of political discussions.

The topics of conversation on Facebook are broad. Users communicate about life events, share content, and there is also quite a bit of political conversation and sharing of political articles [25, 77]. There is more discussion about politics on Facebook than on other social networking sites [25]. In fact, during the 2010 congressional elections, Facebook users could confirm having voted by allowing a badge to be displayed on their page. This created social pressure for other friends to also go out and vote. From this social nudge, Facebook determined that 340,000 Americans who would not have voted otherwise, went to the polls thanks to Facebook [7]. As a political tool, Facebook continues to be at the center of the debate.

3.3 Political context of this study

We conducted this study with participants located in the United States, tracking their exchanges on political events between March and April 2013. Prior to this, on December 14, 2012, a mass murder shooting at the Sandy Hook Elementary School in Newtown, CT, spawned gun law debates. In May 2013, three states, Rhode Island, Delaware, and Minnesota, legalized gay marriage. Our study spans three controversial political events in the beginning of 2013: the federal budget cuts on March 1st, the same-sex marriage debates on March 26th and 27th, and the Senate gun control vote on April 17th. Debates started before, and continued after these dates, but we used them to launch our survey.

	Budget cuts Mar. 1st	Same-sex marriage Mar. 26th	Gun laws debates Apr. 17th	Total
Sample	500	700	700	**1,900**
Responses	19	42	42	**103**
Response rate	3.8%	5.8%	5.8%	**5.4%**
Female	9 (47%)	21 (51%)	26 (62%)	**54%**
Age 18-40	13 (68%)	29 (69%)	24 (57%)	**64%**
Age >40	6 (32%)	13 (31%)	18 (43%)	**36%**

Figure 6: Demographics of survey responders.

3.4 Methods: Survey and interviews

To understand relationship management around political differences, I targeted politically active social media users. As a proxy for users with opinions about U.S. politics, I used the following recruitment criteria: "users who tweet a link to a WhiteHouse.gov petition online." The WhiteHouse.gov website prompts signatories to share the petition link on social media. A user who broadcasts such a link broadens the petition's exposure, and essentially takes a political action.

For each event, I collected tweets containing a link to a WhiteHouse.gov petition, through the Twitter search API, and selected a random sample. I then replied to tweets that were written in English, with a link to my survey. I excluded retweets to focus on personal political viewpoints, and I excluded verified accounts on the off-chance that a celebrity mentioning our survey would induce a snowball sample.

Instead of using Facebook directly, I used Twitter to recruit participants. While somewhat roundabout, tweets are easily searchable, and replying to a tweet is free, which allowed me to reach a large sample. This also allowed me to obtain a range of behaviors on Facebook: those who post about politics on Facebook, and those who withhold from doing so.

3.5 Participants

Via our recruitment process described above, we sent our survey to 1,900 unique Twitter users between March 1, 2013 and April 30, 2013. The survey included a total of 34 questions (see survey in appendix B). We received 103 responses, corresponding to a response rate of 5.4% (see Table 3). About half (54%) of our respondents were female, and two thirds of our participants (64%) were under 40. This is representative of Facebook users [23]. In addition, our participants tended to skew liberal (65% on social issues and 47% on economic issues), also consistent with the demographics of social network sites (see Table 4) [76].

Facebook is indeed a place to talk about politics: 79% of the survey respondents posted at least one thing during the political events. Privacy controls to limit posts to a small audience were rarely used [81]. Participants preferred not to post anything, rather than set a privacy setting.

3.5.1 Follow-up interviews and analysis

Survey participants could choose to provide us with their contact information for a follow-up interview. From this, we interviewed 13 participants. The interview was scheduled for 1 hour, and participants received $30 for their time. I conducted all the interviews over the phone, and participants were encouraged to open Facebook to support recall during the interview.

The interviews were semi-structured, though the script was followed consistently. Questions asked during the interview centered around similar issues as those asked in the survey, but in more depth. In the interview, participants were asked about different relationships that they have with people on Facebook. Furthermore, the interview allowed time to talk about these multiple relationships. I obtained one to three relationships per interviewee, amounting to 33 dyads total.

I conducted an inductive thematic analysis of individual interview questions [4].

Perceived difference	Interpretation	N
0%-29%	High perceived homogeneity	51
30%-59%	Mixed perceived homogeneity	43
60%-100%	Low perceived homogeneity	9

Figure 7: Grouping by how much difference the participants perceived in their Facebook friend network.

To do this, I familiarized myself with the interviews through transcribing and contrasting the interviews with the survey data.

3.6 Findings

A preliminary overview of the participant data shows that most participants perceive their friends to be similar to them (see Figure 7). This parallels homophily research: people tend to associate with people similar to them. In addition, the data suggests that interacting with people of different political opinions is a common experience. The fact that the participants could relate to the experience of dealing with a friend of different opinion could be biased by the fact that we were studying a specific population sample of politically active users. This could be due to homophily (they would be most likely to be friends with others of strong political opinion), or because they were more attuned to noticing posts about politics. Most participants (71%) did not frequently talk about politics with their friends of differing opinion, although many (60%) did see something they did not agree with, and chose not to comment on it.

3.6.1 Tuning out of conversations

Those who perceived a significant number of friends as different from themselves, logged on to Facebook less frequently than those who perceived high homogeneity, $\chi^2(4, N = 103) = 10.05$, $p = 0.04$. With a large number of dissenting friends, Facebook

was less welcoming during political events:

> *I mean it can be overwhelming just being on Facebook [...]. Just like the number of people participating [...], and it was just like there's no room to voice an opinion from either side when I guess your Newsfeed, your friends are just going crazy like that.* (P8)

This could be one reason why political discussions online become an echo chamber, with no one commenting and nobody's mind being changed: those with an overwhelming number of friends with opposing views are not as present. That is not to say these friends never show up. Rather, they are not present when it is most critical, during a debate, both as an active participant and as a viewer. One participant who logged on to Facebook more frequently than usual indicated that this was due to confounding factors like other news happening at the same time. Thus, perceptions of being in the minority directly decreased the desire to log on to Facebook, while perceiving high homogeneity did not have a noticeable effect.

Those who had more friends with perceived different opinions posted fewer comments on their friends' Facebook posts during the political events, than those who perceived higher homogeneity, $\chi^2(2, \mathrm{N} = 103) = 5.61$, $\mathrm{p} = 0.06$. In addition, users who perceived a large number of their friends having different opinions, self-censored more than those who perceived their friends to be similar to them, $\chi^2(2, \mathrm{N} = 103) = 8.37$, $\mathrm{p} = 0.01$. Participating in conversations around political debates often amounted to avoiding confrontation, and siding with like-minded friends.

Rather than joining heated debates, some participants wanted to show support without inviting confrontation. A simple comment or "liking" a post could provide an opportunity to show support in a semi-private setting:

> *I think like one friend wrote [how she felt] I might have commented like I do too, or like word or something like that so more so like I did those*

> *things so that friends [...] know that they're supported by me. But um.. people who are against it I didn't comment back or debate you know with them.* (P6)

One participant mentioned "liking" a friend's posts as a "thumbs up" indicator that does not invite dissenting opinions. The appropriation of these lightweight cues to show support could increase polarization in two ways: 1) explicitly taking an action to side with a party (likely strengthening their own position), and 2) limiting the possibility for someone on the other side to confront them.

In some cases, participants limited their postings to things that brought new insights, or just stayed away all together when there was a noticeable echo chamber effect:

> *I honestly didn't see a whole lot of point in posting anything else related to it because pretty much everyone was in agreement that it was an idiotic thing.* (P2)

> *I like to keep [my friends] on top of what's going on and... what they post I've maybe seen somewhere already so I consider them just an echo chamber.* (P12)

This illustrates that overwhelming agreement can also cause people to hold back from posting on Facebook during political events. Munson et al. [64] found that providing some dissenting opinions in a news aggregator could be engaging for diversity seekers. Having a balance of pro and con posts could help engagement on Facebook during political events. Rather than showing extremes, all supporting or all dissenting posts, people could see a subset of each side of the argument to present a more nuanced display of opinions.

3.6.2 Agreeing to disagree

Facebook arguments were described as long, emotional, and confrontational. Often, they resulted in friends "agreeing to disagree." The following quote illustrates an example of this occurrence:

> *We'll have conversations that reach 80 comments. Um... and like most political debates on Facebook it doesn't just stick to gun control, you go all over the map and then end up back on the same spot where you were and then we just agree to disagree.* (P12)

Many of the survey respondents noted that there was "no point" in engaging with friends of different opinion on Facebook, because they could not "change their mind." These dead-end conversations were described as unproductive and unappealing, meaning that most chose to avoid engaging in them.

The findings show that Facebook is a difficult place to maintain a friendship with someone of opposing views during times of heated political debates. During those times, Facebook can feel like a hostile environment, and the opportunities to show support with like-minded others overshadow the opportunities to engage with people of different opinion.

3.6.3 Weak ties were brittle

Hearing dissenting voices on Facebook created challenges, especially for weak ties. From the participants, I found that weak ties of different opinion communicated less often about politics in general, than strong ties of different opinion, $\chi^2(3, N = 102) = 10.27$, $p = 0.02$. Past research has shown that weak ties are important for getting access to information and being exposed to other ideas [37]. Yet, supporting communication between weak ties is delicate. For many weak ties, Facebook was the only means of communication: 42% of survey respondents who mentioned weak ties communicated with them on Facebook at least once a month, while 4% communicated

by email, 4% saw each other in person, and none talked over the phone. This means that the participants gathered impressions of their weak ties almost solely based on Facebook behavior. Our finding that weak ties primarily communicate through Facebook echoes previous work [34].

When they did communicate, there were challenges. These experiences sometimes made the participants change their opinion about the friends. Other times, it resulted in questioning the relationship, and ultimately disassociating from the friend. Among the 33 dyads mentioned during the interview, three were no longer friends. These relationships were all weak ties based on past common background, such as high school friends who had not seen each other in many years. This confirms findings from [76] that people unfriend weak ties rather than strong ties in light of political disagreements. In contrast, unfriending was not an option with close family:

> *I feel like unfriending my brother on Facebook would cause a lot of strife. Whereas staying friends with him causes frustration but just for me.* (P1)

Simply hiding her brother was a better alternative, because he would not be aware of the fact that she altered the state of their connection. For her, there was still an opportunity for her to stay connected to her brother. For weak ties unfriending each other, Facebook was often the last connection they had.

With weak ties, rather than engaging with their friends, the participants discussed ways in which they actively avoided confrontation. For example, they knew not to talk about politics based on their previous experiences approaching these topics. This meant that they were more careful when talking to their friend, perhaps making explicit choices about topics that should not be discussed:

> *I don't really talk about anything with her other than like small talk and work things that are you know are required.* (P9)

In the interviews, about 30% of the dyads changed their perception of their friend

after the event. From the survey responses, more people felt that their relationship had changed with weak ties than with strong ties $\chi^2(1, N = 102) = 7.68$, $p = 0.0056$. Even though in these cases they may have stayed connected through Facebook, their relationship may have been impacted.

As the medium through which these relationships are expressed and maintained, Facebook has the potential of helping people stay in contact with weak ties. Relationships that are maintained solely through Facebook are particularly vulnerable. Thus, strengthening the discourse happening online could be mutually beneficial.

3.6.4 Tone and feelings of respect

Participants commented on cues they could get from the style of communication of their friend to determine their emotional state:

> *All caps, the frequency that they're posting replies, you know it's usually pretty common for people to post frequently, misspell words, you know no punctuation, all caps, when they're upset they're just typing it together.* (P7)

Often, these perceptions did not match the memories they had of that friend. They commented on discrepancies between the behavior of their friend online and offline:

> *He confused me so much because he was a relatively sort of a hippy guy and um.. I always found him very intelligent and very um.. sort of self educated guy, an intellectual in his own right, um.. and I thought that he was open minded um.. but the way that he expressed himself about the shootings and gun control.. um I found so extreme and so um.. in my mind ignorant that um.. it shocked me.* (P9)

This participant unfriended the person in question, after this one incident. While their difference of opinion may have been sufficient for the participant to disconnect

from this person, the participant points to the means of communication as the ultimate issue. Participants changed perceptions of their friends from the way in which they expressed themselves online, suggesting that encouraging politeness to express differing opinions may have an impact. Turning to Goffman's theory of face [36], these were situations in which saving face on Facebook was a challenge.

3.7 Discussion

Social media might be able to create bridges across ideologies. Insofar as this cross-communication currently occurs, the findings suggest that users may remove themselves from the conversation or from the website. Currently, muffling political discussions, or at least discouraging them, might create a more welcoming environment. This has the consequence of further digging trenches between friends of differing opinions. Designing social media towards nudging users to strengthen relationships with weak ties of different opinions could have beneficial consequences for the platform, for users, and for society.

This matches the behaviors that Pew found. In their survey of 2,000 U.S. adults, they found that these behaviors were stronger for those with strong political beliefs, and liberals more than conservatives [62].

I found that tone and feelings of mutual respect were hard to interpret or understand from Facebook. Designing for supporting civility in social media can have wide-reaching implications in making people communicate across these differences. Social media platforms such as Facebook, which are characterized by real ties, are built around a model that reinforces homophily. Making people of different opinions listen to each other risks threatening the livelihood of this kind of network, since people tune out. I believe that a social media platform designed around the concept of supporting civil exchanges can change the way in which people currently communicate online.

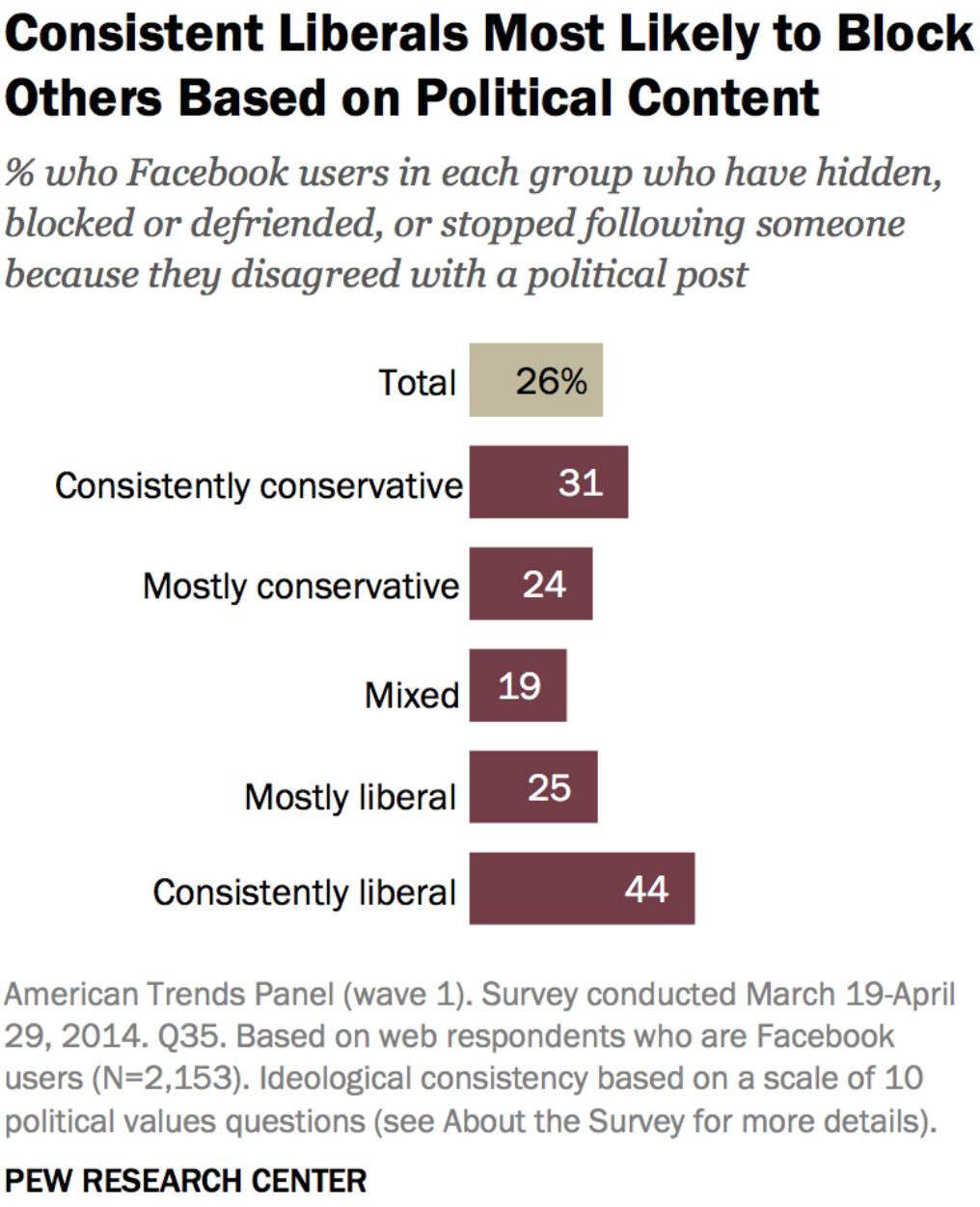

Figure 8: Blocking behaviors on Facebook reported in [62].

Bursting the filter bubble can happen in existing social networks, simply by hearing the voices of those who think differently. Our data shows that people with many friends of opposing views decrease their activity on Facebook during political events. Their absence from the debate means missed potential for discussion and deliberation. This reinforces the echo chamber. Not only are people mainly friends with like-minded others, but their few friends of different opinions will not expose them to other views. On Facebook's Newsfeed, where the algorithm for displaying posts is unknown, this effect might be redoubled: people who do not communicate with each other do not appear in each other's Newsfeed.

Social streams, such as the Facebook Newsfeed, would be interesting places to experiment with modulating the amount of posts shown about a topic. We saw that some of our participants decreased their use of Facebook during political events because it was overwhelming to log on. Facebook does not give an option to limit topics in the Newsfeed. The only controls available are limiting specific individuals from appearing though hiding them. Limiting topics seemed to affect the ads shown, in addition to the Newsfeed. This was mentioned by one of our participants as a turn off: he noted an increase in ads about guns during the debates on gun control. Overall, these topical controls may provide a more pleasing experience during controversial political times, as well as other "hot" events.

This last point suggests something similar to what Munson et al. [64] found about providing dissenting opinions in a news aggregator. Having a balance of pros and cons could be helpful for making Facebook seem like a more interesting and inviting place during heated political events. Rather than showing extremes by displaying all dissenting posts, the Newsfeed could select a subset of each side of the argument to present a more nuanced display of opinions. We speculate that these changes to the Newsfeed may bring more weak ties to light since they are the ones who are currently less emphasized in the Newsfeed. Hopefully the fact that the Newsfeed

would be more balanced would help, but we saw evidence that weak ties are brittle to dissenting opinions, so this needs to be tested.

3.8 Implications

My initial question was: can we design social media in a way to make people more civil? Underlying this question is trying to understand whether social media plays a role in increased polarization, and as such, is social media inherently polarizing, or is it a matter of choosing certain design paradigms to change this?

These opportunities posit that social media could better facilitate discussions across ideologies. Some suggestions resulting from this work include calling attention to past interactions and shared interests, to make common ground visible during arguments; or creating data visualizations and interactive statistics to support more civil conversations grounded in objective data. Ideas around slowing down contentious discussions to make weak ties more resilient such as in [20], also merit consideration. In the rest of my dissertation work I attempt to understand design implications to facilitate these discussions.

One of the largest and most consequential behaviors we saw, which was tuning out because of too many people "going crazy" in political arguments on Facebook, could be addressed by weighing posts that appear in the Newsfeed to varying degrees. For example, if we can detect the degree to which someone might agree or disagree with a political post, similar to what has been done on blogs [2], then this could be taken into account in order to post a balance of pro and con posts. However, this could also have the consequence of increasing the echo chamber effect rather than bursting it. If people are given the option of calibrating the algorithm, then those who are more "challenge adverse" [64] may limit their exposure to non-agreeable others.

Along these lines, I suggest a social interaction design to nudge people towards behaviors that make it easier to discuss across differences. Simple design elements

can create norms in social networks. For example, the "like" button on Facebook has a certain social etiquette around how it should be used. In this study I found that people used the like button in political conversations to show support towards a statement from a friend, without inviting other comments or engaging in the conversation. Facebook has held off including a "dislike" button because of potential backlash and abuse.

In the case of heated discussions on social media, perhaps some positive visual cues might help to draw in friends who are particularly pro-social or positive conversationalists. This way, hearing from friends of a different opinion is still possible, and might be a visual representation that someone is at least trying to not harm or hurt others. Similarly, removing heated comments from the platform could be an alternative mechanism. Blocking and removing deviant individuals from forums is a common community moderation guideline, and Facebook users are well familiar with the potential to unfriend or block a dissenting friend. However, completely removing a friend from a social circle might be overkill, and could be avoided by a curation of only their posts that meet a certain threshold of civility. Studying the components of a platform designed in this way would shed light on designing social media platforms to alleviate the most pressing issues brought out in this study.

The question is the following: can we design social media in a way to make people more civil? How do we choose certain design paradigms to change conversational tone in social media? Exploring these questions made me consider the ultimate challenging question faced by any social computing researcher: Do I need to design my own version of Facebook? Studying alternative designs in social media ultimately raises this daunting question. Building the platform itself is not the issue, but obtaining the social graph required in research is akin to having built a company like Facebook from scratch. In the next Chapter, I describe a prototyping technique I utilized to build a browser plugin that modifies the UI of Facebook, rather than build a social

network from the ground up. This method can be used in other contexts as well, as my examples show. The results from the deployment of my prototype to bring about a more civil experience on Facebook is described in Chapter 5.

CHAPTER IV

PROTOTYPING SOCIAL SYSTEMS

Understanding how to build pro-social systems and empower people with them requires that we on the research side are able to build and deploy these systems. Because of the near impossibility of doing this, we do not necessarily build complete production systems, but rather artifacts that are continually iterated on and known to have flaws or be partially complete.

What is especially difficult is prototyping social systems, because of the complexities involved in gaining critical mass. If we wanted to evaluate ways of designing social media to be more civil, we would need to start by building a social media platform. How could we understand the implications around civility, unless a large group of friends, and friends-of-friends at the very least, had opted-in? The uncertainty involved with scaffolding a social app to this scale could have stalled our project. Even massively funded and pioneering commercial systems (such as Google Wave and Color) have failed to achieve this critical mass of users. Thus, building a social media platform was risky: it would cost significant engineering resources, and what if no one used it? Did we have to build the system end-to-end just to evaluate our design ideas?

Of course, this is what prototyping is designed to solve. Prototyping can pinpoint fundamental flaws in interactive systems before a design team invests considerable energy building the system. Most HCI systems start as prototypes: they are designed and developed iteratively, and at increasing levels of fidelity. Unfortunately, existing HCI prototyping techniques do not translate well to social computing systems. Rather, they focus on prototyping *interface interactions*, rather than *social*

interactions. Interface interactions refer to an interest, and focus on interactions between a person and elements of a technological artifact, such as the layout of a website for a usable experience. I distinguish this from interactions between people (one-to-one, one-to-group, group-to-group) mediated through technology, where the primary interest is facilitating pro-social behaviors (connecting strong ties, meeting new people, knowledge communities, moderating deviant behavior). Usability may still be a consideration in social systems, and similarly, social interactions may appear in projects focused on usability, yet one must be emphasized over the other. Methods for prototyping interface interactions are well developed in HCI, while we are greatly lacking frameworks and methods for studying social interactions.

My solution to this shortcoming was to develop a novel prototype technique, which I call piggyback prototyping [39] [1]. It is a 6-stage prototyping mechanism for testing and iterating on new social computing designs, and works by coupling semi-autonomous bots or plugins to existing successful large-scale social computing systems. Piggyback prototyping overcomes the challenges of obtaining critical mass, by leveraging existing social platforms, such as Twitter or Facebook. This allows researchers to focus on what people do on their system, rather than how to attract people to it. A piggyback prototype focuses on exploring social interactions, rather than interface interactions. Specifically, social interactions are those involving people directly exchanging with each other. Such exchanges can involve messaging or commenting, liking or upvoting, meeting face-to-face, sharing information, etc. We are not using "the crowd" to add information, or change the display of an interface. Rather, Piggyback Prototyping is used for systems in which people directly communicate with one another.

[1] This Chapter extends work published as C. Grevet and E. Gilbert. 2015. Piggyback Prototyping: Using Existing, Large-Scale Social Computing Systems to Prototype New Ones. CHI. *Honorable paper mention*

The contributions of this Chapter are threefold:

- First, I will describe the steps of piggyback prototyping, illustrating with examples along the way.
- Second, I will describe my experience using a piggyback prototype in a real-life project, which consisted of creating a Chrome plugin that hides impolite posts from the Newsfeed, and highlights polite posts.
- Finally, I will discuss the scope, benefits, and limitations of the piggyback prototyping technique.

4.1 Piggyback Prototyping

Piggyback prototyping is a social computing system prototyping technique that utilizes existing social platforms to evaluate novel social interactions for large-scale systems. It is best suited to projects that require coordination between multiple people, and that do not have access to readily available large-scale social data. Many social matching systems could have been prototyped with this technique, such as organizing people for disaster relief, assisting a collaborative activity, matching people according to interests, and others. Piggyback prototypes involve 6 stages, two of which distinguish it markedly from other techniques: a non-social pilot (stage 3 in Figure 9) and a social deployment (stage 4 in Figure 9). The evaluation of the prototype will likely involve mixed methods in which a researcher might craft a survey, plan an interview, collect log data, or compile user responses.

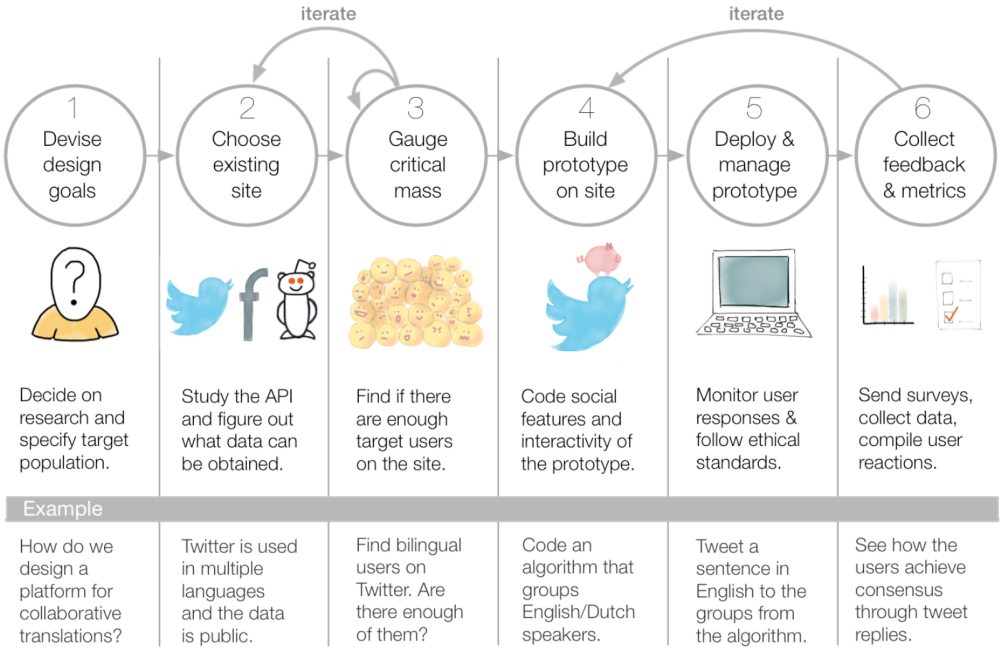

Figure 9: Piggyback prototyping

4.1.1 Devise design goals

The first step is to decide on the research goals and the desired social interactions. Figure 9 presents an example for a collaborative translation system that resembles the existing site Duolingo [2]. In this step we also suggest determining the target population for the prototype. Identifying these elements will be the basis for the next step, and for planning the architecture of the prototype.

4.1.2 Chose existing site

There are many existing social platforms that could be used in piggyback prototyping, for example, Twitter, Facebook or Reddit. The researcher should consider the pros and cons of the different platforms. We recommend choosing a platform with public data and an API. In particular, three types of data shared on these platforms may have value to designers [42]:

[2] duolingo

People: User profiles contain a wealth of information, such as demographics, interests, and network data. If it is possible to obtain a social graph, then systems requiring this structure could be prototyped. In that case, the prototype might require some sort of massive scale snowball sampling [5], to get friends of friends to participate.

Things: The things that people talk about create vast amounts of content around current events, interests, and sentiments. A researcher could build a prototype that centers on topics. There exist off-the-shelf natural language process packages (such as NLTK4), and topic models, that can provide enough accuracy to develop quite sophisticated tools.

Places: There are a number of sites where people broadcast their location (Twitter, Foursquare). These platforms may serve different purposes, and motivations for publicly sharing geo-location information may be complex [15, 85, 56]. Understanding these intricacies will help align a prototype with user expectations.

4.1.3 Gauge critical mass

Finding critical mass on the chosen site for the target population is key. Critical mass is not an absolute value, but rather is specific to each project. If the goal is to match people based on topics, then enough people need to be talking about a given topic. We suggest gauging critical mass on the chosen site by testing the participant recruitment method without a social component. This pilot can be a simple message sent to users, or it may be a survey. This can help determine characteristics of the target population such as: demographics, behavioral patterns of the target population, and technical aspects such as the types of devices typically used by the participants.

As described in Figure 9, gauging critical mass is an iterative process. Using our experience as an example, in Chapter 3 I tried recruiting participants for a study about political discussions on Facebook. I initially planned to recruit participants

who shared a link to a political petition in a public Facebook post. I quickly observed hostility towards our requests via Facebook, which was not the case when we reached out on Twitter about the same topics. The norms and expectations were different, and I learned that Twitter was a better choice for that study. This example does not suggest that Facebook is an impractical site for piggyback prototyping. It may be just right for certain systems, and in fact is the platform I chose to use for my study in Chapter 5.

4.1.4 Build prototype on site

Once the site is chosen and proves to have enough potential participants for a viable study, the social aspects of the prototype can be built. Piggyback prototypes might be semi-autonomous: running code may find users and send them specific messages. Some prototypes might suggest that participants communicate with each other through the existing site. In the collaborative translation example from Figure 9, groups would communicate through Twitter replies. Other prototypes might ask participants to communicate through a low-fidelity interface. For example, Facebook users might be asked to collaborate with others on a simple and easily deployed message board.

Broadly, there are two components that are well suited to be tested through a piggyback prototype:

Algorithms: piggyback prototyping can test social algorithms. Recommender algorithms, and natural language processing, are examples of such algorithms that are commonly used in social computing, and that are particularly difficult to evaluate without critical mass. Since a prototype is rough, these tools need not be perfect. In fact, piggyback prototyping could help evaluate them in a live setting.

Interaction: this concerns the messages that are sent to users, the tools available

to users through the prototype, and the means by which users communicate with each other. Piggyback prototyping can help iterate on these aspects.

4.1.5 Deploy and manage prototype

Next, the researcher or designer deploys the prototype on the chosen site. Systems prototyped with this technique should strictly adhere to Internal Review Board (IRB) processes, or corporate ethics boards if applicable, and always to ethical standards. Prototypes should not violate privacy, and should only present minimal risk.

4.1.5.1 Obtaining consent

We worked closely with the IRB at our institution to conduct our study using piggyback prototyping (presented next). We highly recommend doing so. Some projects may suffice with a waiver of consent, without need of documentation, when they present minimal risk. Piggyback prototypes can work with documentation of consent as well, as long as the necessary critical mass of participants signs it. Full-disclosure of the research goals and obtaining consent may be prohibitive for some projects. Those studies fall outside the scope of piggyback prototyping.

4.1.5.2 The role of the researcher and self-presentation

Even though piggyback prototyping involves running code, it is not completely autonomous: the designer/researcher is still a central part of the prototype. They must present themselves as such to the participants with whom they interact. Moreover, similar to a wizard-of-oz prototype, in which the researcher is part of the system, in piggyback prototyping the researcher must be deeply involved in the process. This means that the researcher needs to be available to conduct duties such as answer specific participant questions if appropriate, or to remove participants who have asked to be excluded, or who behaved poorly.

4.1.6 Collect metrics and feedback

The goal of prototyping in HCI is to evaluate a system in order to iterate on the design [3]. Participants of a piggyback prototype can be sent a follow-up survey to ask about their experience. They might even be interviewed, although the number of participants in this prototyping technique might get overwhelming.

We propose that the following metrics can be obtained through piggyback prototyping:

4.1.6.1 Engagement metrics

These are the data that can be obtained from the social platform and supporting ecosystem. For example, the number of clicks on the supporting documentation can serve as one indication of how many people saw the message.

4.1.6.2 Survey evaluations

The gauging critical mass survey and final survey are two entry points to ask users for their thoughts on the system. These surveys could ask usability questions about their interactions, or could also ask behavioral questions.

4.2 Examples amenable to Piggyback Prototypes

Here we will present three short examples of existing systems that we believe could have been prototyped with piggyback prototyping, yielding informative research directions. These examples are commercial systems that have already proven their success as viable commercial products. While we now know how users interact via these sites, it is an interesting exercise to apply piggyback prototyping to them.

4.2.1 Online dating: Can we prototype OkCupid?

Piggyback prototyping could help evaluate certain components of an online dating site such as the matching algorithm. What algorithm results in the most relationships

being formed? Who are people looking to meet? For example, we could find people who tweet about being single, and about a sports team. Do those who root for the same teams end up getting along? The evaluation could ask whether they would be interested in seeing that person again.

4.2.2 Expert knowledge systems: Can we prototype Quora?

We could imagine studying a knowledge system across many people. For example, we could find people interested in the same topic in subreddits, and ask them to contribute to a shared document. Studying the design of this system has many important research contributions. Who contributes? What topics make the most sense for this? What incentives foster the best answers?

4.2.3 Co-location meet-ups: Can we prototype Foursquare?

People go about their daily lives in public places where they are in the presence of others with similar patterns, known as familiar strangers [61]. We could try to increase the social capital present in a city by pairing people who tend to tweet from the same location. How should these meetings occur? Does this indeed increase social capital?

4.3 Building a plugin for civility on Facebook

In my instantiation of piggyback prototyping, I developed and evaluated a plugin for Facebook that aims to create a more civil social environment. In the rest of this Chapter, I provide the design decisions and engineering undertakings for this plugin, using the piggyback prototyping framework. In the following Chapter, I will describe the study setup and evaluation. The prototype is a plugin for Chrome that modifies the Newsfeed on Facebook in terms of civility, using a politeness classifier.

4.3.1 Our prototype goal: What would make Facebook more civil?

The motivation and design decisions for this prototype stem directly from the findings in Chapter 3: in the case of heated discussions in social media, perhaps some positive visual cues could help bring to the surface conversationalists who are particularly pro-social or positive in their tone. This way, hearing from friends of a different opinion is still possible, and it introduces a visual representation of constructive contributors to the conversation, even when there is disagreement. Similarly, removing heated comments from the platform is an alternative approach. Blocking and removing problematic individuals from forums has been a common community moderation guideline, and Facebook users are well familiar with the potential to unfriend or block a dissenting friend. However, completely removing a friend from a social circle might be overkill. Instead, selecting only their posts that meet a certain threshold of civility would be kept.

4.3.1.1 Main components of the civility plugin

Hiding impolite posts and comments: II found in the study in Chapter 3, that people tune out of Facebook during politically controversial times, because "everyone is going at it". So my plan was to build a plugin that would hide contentious comments. A user who installs my plugin will not see all the back-and-forth. When he logs on, he may see some posts and discussions he disagrees with, but it will limit the quantity of these posts.

A participant in my previous study was frustrated about her brother's political posts. Her only option was to hide him from her Newsfeed, meaning that she never saw his posts. This plugin will remove entire posts that are deemed impolite. If this user installs my plugin, it will hide the condescending posts from her brother. She will still be able to see the pictures he posts of her nieces.

Highlighting polite posts and comments: I also found that people changed their perceptions of their friends, particularly regarding weak ties. Highlighting the civility of friends might help to humanize someone with very different views. Through highlighting civility, a user can notice a friend's ability to be thoughtful, despite having different views. With this plugin install, a user who logs into Facebook will see polite posts and comments in their Newsfeed as highlighted in green.

Politeness feedback about status updates: Participants commented that they would get drawn into long pointless conversations on Facebook. A user who installs my plugin will get feedback on a message he is about to post if it is classified as impolite. After being alerted that his message is impolite, he is more likely to think twice about it, double check his facts and soften his language. The following reply in the thread appreciates the thoughtfulness in his comment.

4.3.2 Our selected existing social network: Facebook

I built this plugin on top of Facebook as an existing large social network. I found in my formative study that there is a problem with incivility on Facebook. It is the biggest social media network, and so has significant implications in relationships between people - also it contributes to the filter bubble.

4.3.3 Did we find critical mass on Chrome/Facebook? Yes

Part of this finding was informed by my prior study on how people maintain relationships with friends who are different from them on Facebook [40]. It brought issues to the surface that were faced by many participants regarding heated conversations. For this prototype in particular, I needed enough Chrome and Facebook users to make it work. I sent out a screener to find participants for the study, asking what browser they use, how often they use Facebook, and on what platform they primarily

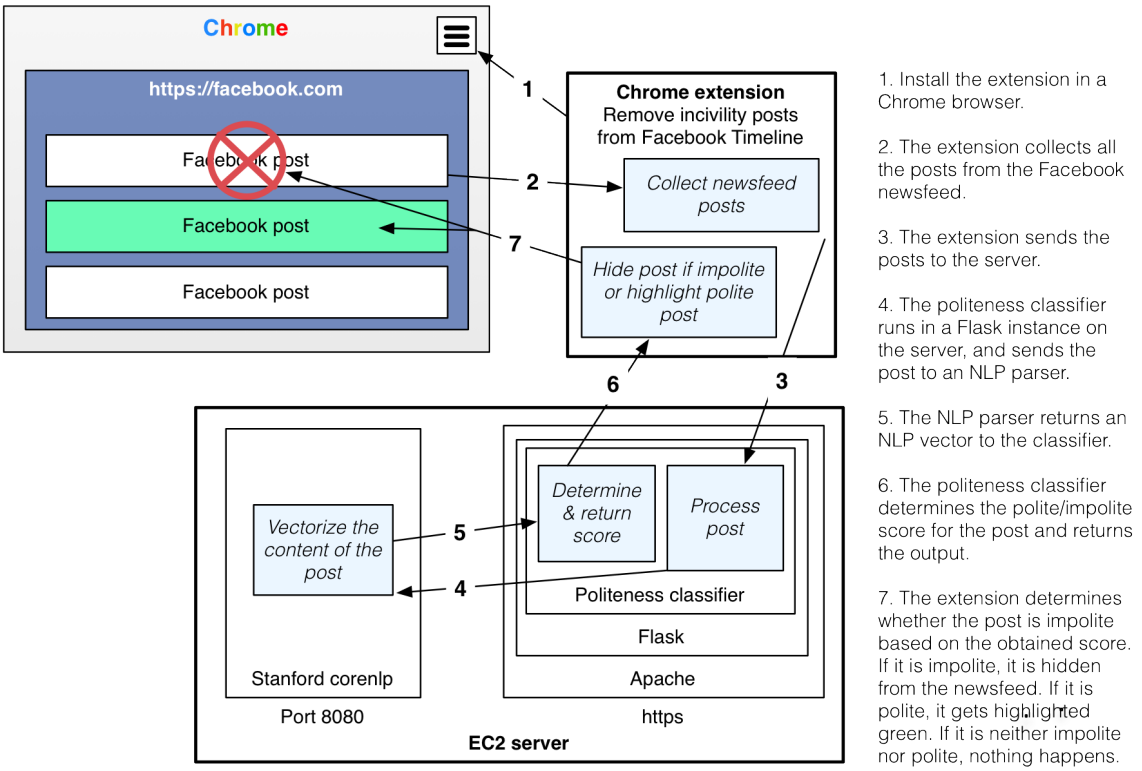

Figure 10: The technical architecture of the civility prototype.

use it. I found critical mass in my target population of frequent Facebook users on Chrome, and thus knew to build a Chrome plugin. However, there were not enough participants who primarily used it from their Desktop computer. Since that did not present a technical roadblock, I continued with this sample population.

4.3.4 Our prototype design & algorithm

My prototype for Facebook contains two main components: a browser plugin (built as a Chrome extension), and a server-side politeness classifier (see Figure 10). I will describe the technical details of these tools.

4.3.4.1 Prototyping in tampermonkey

Tampermonkey[3] is a script manager for Chrome that facilitates creating scripts. It allowed testing the feasibility of the front-end aspect of this plugin before building a

[3]https://tampermonkey.net/

complete plugin. Leaving the plugin as a Tampermonkey tool can be done in certain types of studies (such as when participants come to the lab to test a client-side interface). Yet, for the purpose of a diary study deployment on participants' own machines, it was much easier to deploy as a stand-alone Chrome extension.

4.3.4.2 Chrome extension

The Chrome extension was deployed as a standalone, and was sent via email to the participants. Recipients were walked through the process to install the plugin in their browser through developer mode, rather than through the Chrome Store, in order to limit the spread of the plugin. Each participant was also given a personal 4 digit pin to log in to the plugin. This pin was attached with the survey responses of the participant, as well as with the log of the plugin.

4.3.4.3 Collect Newsfeed posts

Once the plugin is installed on a Chrome browser, it listens for activity on the Facebook Newsfeed page. When the page loads post to the Newsfeed, the civility extension collects them and sends them over the server to get classified. Two precautions were taken to ensure the safety of our participants in this process: 1) only the content of the post was sent to the server (not the name of the author of the post), thus posts were sent anonymously, 2) the communication was encrypted over https, which is required of Chrome extensions that communicate with external servers. The server was upgraded with an SSL (Secure Sockets Layer) certificate.

4.3.4.4 Send posts to classifier

The server was built in an Amazon EC2 instance running Apache (version 2.4.7). On the Apache server, we deployed the web application framework Flask[4] (version 0.10.1) to run the politeness classifier. Flask then queued incoming requests to a Celery [5]

[4]fhttp://flask.pocoo.org/
[5]http://docs.celeryproject.org/en/master/index.html

queue. The queued requests were sent to Stanford CoreNLP [6], running on a separate port. Stanford CoreNLP parses the incoming Facebook message to a set of linguistic parameters, that are then analyzed by the politeness classifier as described in [16].

4.3.4.5 Receive output from classifier

The classifier determines a politeness and impoliteness score for each message. Each score is ranked from 0 to 1, and the sum of both scores equals 1. When this score is returned to the Flask web application, the scores for politeness and impoliteness are compared against a provided threshold. In this current instantiation of the plugin, the threshold for politeness was set at 0.65, and the threshold for impoliteness was set at 0.60. This was based on an evaluation amongst members of the lab, and I found that while we tended to agree on politeness for higher levels of politeness, impoliteness was less accurate at higher than 0.60.

4.3.4.6 Rendering Facebook Newsfeed posts

The web application would then return the score, along with the determination about whether it was considered polite or impolite. From this, the Chrome extension would determine how to render that particular post on the Facebook Newsfeed. If the post was considered impolite, it would remove the entire post, including all attached comments, from the stream of posts in the Newsfeed. If it was a comment that was considered impolite, then the comment would disappear from the comment thread. If a post was considered polite, then that post, not the associated comments, would be colored in green. If it was a comment that was considered polite, then that comment would appear in green.

Because this processing must occur after the Facebook Newsfeed has been loaded (in this case we listened to new scroll events to update the rendering of posts as a user interacted with Facebook), and because there are many layers to this prototype (for

[6] http://nlp.stanford.edu/software/corenlp.shtml

example, Stanford CoreNLP could take a while to parse text), at times there appeared to be a lag in how it rendered posts in the Newsfeed. Some of the participants commented on posts disappearing while they were reading them.

4.3.4.7 Diary survey

The classified posts, aside from being rendered based on whether they are polite or impolite, were also stored locally in the storage of the participant's Chrome browser. Chrome extensions have access to sandboxed storage, which allowed us to store the posts that were classified during the day, and show them at a later time to the participant. At around 8:00 p.m. at night, participants received an email to complete their daily survey. While the link in this email worked at any time, there was a link displayed on Facebook as well, between 8:00 p.m. and mindnight, to encourage them to complete it then. The survey will be described in further detail in the study setup in Chapter 5.

4.4 Other instantiations of piggyback prototyping

In addition to the example described above, in a separate study, I also used piggyback prototyping as a framework to develop a prototype for introducing strangers who are different from each other in airports. Along the same thread as designing social media to allow people to hear from others of differing backgrounds, I carried out a project aiming to help them meet face to face. For this prototype, I used Twitter as an existing site, and sent pairing messages to Twitter profiles of people who had sent a tweet from an airport.

4.4.1 Prototype goal: Will people meet strangers?

The goal of this prototype was to nudge people to meet others different from themselves, in order to combat our natural tendency towards homophily [60]. Previous

research on Political Blend, a system designed to introduce people of different political beliefs over coffee [22], showed indications that such meetings could be valuable. Yet, prototyping the system itself was challenging. Through piggyback prototyping, we can answer the following questions:

- Are people willing to meet strangers when prompted through social media?
- Are people willing to meet strangers despite not having much in common?

4.4.2 Our selected existing social network: Twitter

We chose Twitter as an existing large social network. The Twitter API provided us with the ability to obtain public data, including user location, social networks, and profile data. We chose U.S. airports as locations for introducing people, since they have benefits such as: many diverse individuals might be collocated; engaging in a conversation with a stranger may be a pleasant way to pass time; and importantly, airports have significant security procedures that may lower risk in meeting a stranger. Furthermore, airport check-ins on Twitter seemed like a viable route, following the success of the TSATracker system, which asks Twitter users for updates about security lines [69].

We ended up forming 3,161 pairs, from which we received 576 tweet replies, 183 survey responses, and 8 participants who actually met in person. We learned that people would, in fact, meet others through our envisioned system, and more importantly, specifically how future design iterations could facilitate more meet-ups to occur.

4.4.3 Did we find critical mass for airport check-ins? Yes

We first conducted a formative survey for three purposes: 1) to determine whether people would be willing to meet strangers in airports; 2) to help us determine an expected response rate using airport check-ins on Twitter; and 3) to give us insights

into the demographics of this population. The survey asked whether they had met a stranger that day in the airport, whether they would use an app to meet strangers in the airport, as well as demographic questions to explore the diversity of the population.

We sent surveys to 1,512 Twitter users who checked in to a U.S. airport between December 2013 and Jan 2014. They received a request to fill out a survey about an app to introduce people in airports. We obtained 213 responses, with a response rate of about 14%. From their responses, we found that our target population was interested in meeting fellow travelers, and that there was critical mass for users who checked-in within 15 minutes of each other. This was encouraging to pursue further down the road.

4.4.4 Our prototype design

Once we had early evidence that some significant groups of people were interested in meeting strangers in airports, and that the study was feasible, we prototyped the actual interaction involved with matching users and prompting them to meetup through Twitter.

For this prototype, we paired users based on their similarities and differences. We built a Twitter similarity classifier that is based on common known dimensions of homophily [18]: content of a user's tweets, their followers, and who they follow. We assigned our participant pairs to three user groups (high similarity, some similarity, low similarity) based on the obtained similarity score. These thresholds were obtained from formative data collection resulting from determining critical mass. The implementation of this algorithm was done in Python with the Tweepy[7] module to connect to the Twitter API.

Through the Twitter Search API we searched for users who checked in to U.S.

[7] tweepy

airports, then for each user we obtained their social graph and their last 20 tweets. We created clusters of users per airport who checked in within the last 15 minutes. Within each cluster we computed the similarity score for each possible pair. We sent a matching tweet to those who were 1) most different, 2) most similar, and 3) finally, if some pairs were left, they were paired up. Our prototype consisted of approximately 1,000 lines of code, of which a significant proportion was the similarity algorithm.

4.4.5 Prototype deployment & lessons learned

This prototype ran on weekdays from May to September 2014. We did not run it continuously during that period, since we did not know what to expect from user responses, or from Twitter where our account could have been blocked. We started by sending tweets manually, and eventually turned to a semi-automatic system when we found that most responses were positive. We stopped contacting those who requested it (such as not sending them our survey). This only concerned 10 pairs of the 3,161. All the follow-up surveys were sent manually. By monitoring our study closely, we could iterate on some aspects of our prototype:

First, we initially contacted participants as the primary author. This put all the strain on the researcher's account since it meant contacting thousands of users. For that reason, we changed to a more general research account. Our contact information was still available in the study documentation.

Second, we iterated on the message crafted for introductions [86]. Each pair (e.g. A and B) was sent the following prompt: *"@A and @B you're both at CLT. Why don't you meet before your planes take-off?"* The prompt was followed by a tweet containing the age limit of 18 for participation, and another tweet with information about the study. We got feedback that the tweet about having to be 18 years old was deterring, so we included it in the same tweet as the information link. This streamlined the process by only sending two tweets (one to initiate the match, and

one with the study information).

4.4.5.1 Evaluation: survey instruments and other data

The day after we paired Twitter users, we sent them a link to a survey to ask them about their experience with the meet-up (see Figure 3 and Figure 4). In this survey, we asked them whether or not they did meet up. If so, we asked them questions on these three topics: inter-personal likeability [43], self-disclosure, and followup connections. Finally, we asked them how long the meet up lasted. In the cases where the participants were not able to meet up, they were asked why. We asked all participants if they would like to be matched again. Most of the questions were asked as a Likert scale.

In addition to the surveys, we also obtained data for our prototype through page views analytics and the tweets we received back from our participants. This data was simply obtained from a Google Analytics script inserted on the documentation page residing on our lab server.

4.4.5.2 Lessons learned from this prototype

The goal of our prototype was to see if people would meet face-to-face, and to gain design insight into a system that would prompt people to do so. We did see people meet, and the engagement we got with the prototype was enough for us to develop insights into the design of a system in this context.

People were willing to meet strangers in airports: We sent a survey to 1,512 Twitter users who checked in to a U.S. airport between Dec. 2013 and Jan. 2014, and we obtained 213 responses. In this survey, we asked about whether they had talked to a stranger since they had been in the airport. Over half of the users who checked-in to an airport on Twitter had engaged in a conversation at the airport with someone they did not already know (56%). When asked if they would be interested in meeting strangers while they waited, 71 participants (32%) said "yes" and 112

participants (51%) said "maybe." Only 33 participants (15%) said that they would not be interested in meeting strangers. Of those who said they would not be interested in meeting someone while they waited, there were only 10 of them (21%) who would not use a social app in the airport. Others might install an app to get a coupon with someone while they wait (35%), or to play a social game (20%). These were key findings that allowed us to move forward.

Participants engaged with our prototype: From May to September 2014, we paired up 6,322 Twitter users (3,161 pairs) who had checked in to airports on Twitter, and sent a follow-up survey the next day. We obtained 186 survey responses, of which 182 respondents had not met their match, and 4 had met their match. We got 576 Twitter replies and we had 712 unique visitors to our study information page. Our pairing tweets were favorited 61 times. Of the replies we received, 31 had location or contact information. These data suggest that a rough social prototype like the one we deployed can lead to significant amounts of data that help gain insights on the intended interaction. The replied tweets can be analyzed to understand what is happening. The fact that the tweet was repeatedly favorited (61 times) is encouraging. And we obtained enough survey responses to gain a more in-depth understanding, which we will talk about next. An example exchange between participants is illustrated in Figure 4. Other tweet replies we received were:

"would've made my afternoon"

"haha safe travels! Hope you're not #theOnethatGotAway"

"I hope you're doing this for awhile it's such a cool idea!!"

"I would have participated had I not been so preoccupied with getting my luggage and search for tacos. Next time"

Missed connections: In the deployment, we saw a number of missed connections. 31 participants tweeted their gate or location information as a follow-up to our pairing tweet. We found that the reasons meetings did not occur were: participants saw our

tweet too late; the participants were too far away from each other; the participants did not have enough time before their flight; social reason (such as traveling with a family member); some checked in to the airport even though they were not traveling; and finally some had arrived at the airport rather than waiting to depart.

Actual meetings: What is remarkable is that people actually did meet when prompted through our prototype. In total, eight participants met thanks to our prototype. These meetings occurred between people who were highly different (according to our computed similarity score). Of the four reported meet-ups, the participants felt rather neutral about whether they could be friends with the person. This could be simply because a friendship needs more time to develop. Yet, most participants would be interested in being paired again (3 of the 4 pairs). Most meetings exchanged contact information (3 of the 4). One of the pairs, in fact one of the most different pairs, talked about topics that tend to be more controversial, such as religion and politics. In all cases, they talked about their jobs. Family and relationships, and general interests such as music and movies were also talked about in three of the meet-ups. The participants also reported that the meetings lasted 60 minutes in two cases, 30 minutes for one and 5 minutes for another.

4.4.5.3 Future design considerations

The goal of prototyping is to gain design insight about what would happen when a full system is built and deployed. From our instantiation of this prototype, we learned:

Opt-in system: Most missed connections might not happen if users initially signed up for the system and were thus expecting these prompts. The challenge with building an opt-in system is obtaining critical mass, this was the reason we did not go that route in the first place. Now that we have an idea that people are interested in meeting others while waiting in airports, we can build a complete system with more confidence about our design decisions. In an opt-in system, we would have greater

control in creating pairs that can meet. For example, we could take into account how much time travelers have until they board, and whether they are arriving or departing. Through this prototype, we were able to determine some of these issues that could be explicitly designed for in our next iteration.

Matching message: It was surprising to us that participants did not seem to need to know much about the other person to be willing to meet. While it might have been more motivating to know some common interests with the match, we did not see the lack of information about the match to be a large barrier. Perhaps a finding here is that people do not really care to look at the profile of the other person, and are actually generally willing to meet a stranger, no matter who that person is. A controlled comparative study using piggyback prototyping could more fully explore this. A priori, we thought the introductory message would be key, but it did not seem to matter much in our prototype.

4.5 Considerations with Piggyback Prototyping

As a prototype technique, a piggyback prototype is not meant to be a fully completed system. Rather it is rough and flexible: it should be easy to iterate on. As we described, the piggyback prototyping technique is a 6-stage process that provides a scaffolding mechanism of an iterative process for designing large-scale social computing systems. In our instantiation of piggyback prototyping, we learned about ideas that would improve our initial system, like having it be opt-in for pairing people according to more fine-grained information. We hope to have shown how other researchers can also implement this approach.

4.5.1 Critical mass

Obtaining critical mass in any system is extremely complex and not well understood. Users might come because of good design, a well-timed product launch, or simply

because of good luck. In our prototype, we knew from our formative survey and data collection, that there were enough people checked-in at the same airport at the same time to pair them up. This step is necessary to make sure that a prototype will have enough users. It is not because a prototype is successful that the resulting completed system will obtain critical mass. However, through this technique we hope to give researchers and designers more tools to consider projects that they might not have had the resources to begin otherwise.

4.5.2 Volume of users

Piggyback prototyping concerns large numbers of users. This is unique to this prototyping technique, compared to others used in HCI. As such, the evaluation of a piggyback prototype must be catered to this volume. We would argue that a quantifiable survey is more manageable than user interviews. This also means that the resources to manage the volume of participants must be considered. Participants may want information about the study, or may personally message the researcher. While we only had two cases of participants emailing us for more information, we can imagine that this could quickly become difficult to manage if every participant had emailed us.

4.5.3 Choosing appropriate metrics

Our piggyback prototype made us reconsider the traditional evaluation metrics of social computing systems. We had many participants, due to the fact that Twitter had large numbers of people checking in to airports. Thus, the fact that we had 6,322 users does not speak to the merit of our system. What does? We looked more deeply at survey results, engagement metrics and user responses to get a sense for the value of our system. Similarly, researchers who employ piggyback prototyping should determine for their project what metrics and feedback they would like to obtain.

What was important to us was to determine whether some people would meet

up and whether those meet-ups were meaningful. Some might consider the four meet-ups we saw to be a limiting aspect of our study: "four is a small number, so the impact of the system is underwhelming." Yet, we saw that those meet-ups were highly successful from the survey responses, despite the fact that the people paired up were highly different. This finding is surprising and significant enough to continue down this line of work.

4.5.4 Longitudinal studies using piggyback prototyping

There is a dilemma around the longitudinal aspect of piggyback prototyping. On the one hand, a script could constantly run to obtain data over a long period of time, as long as one does not get blocked from crawling the site. (That is, a site could interpret high levels of activity against it as an attack and shut the script down.) At the same time, one must consider possible user fatigue. To our knowledge, Twitter users who tag their location are not constantly bombarded with research requests. While this study shows that at the time of the study, a significant number of people welcomed our intervention, this could also be due to a novelty effect. If these requests were a more frequent occurrence, Twitter user behaviors may change. While we see promise in the feasibility of this technique, we are also aware that an over-abundance of piggyback prototyping might drastically change behavior, and therefore the feasibility of the technique. This kind of reflexivity is present in most social systems.

4.5.5 Generalizing outside of Facebook

Our piggyback prototype was deployed on Facebook. This platform was ideal at the time of this study because it contained a large public dataset of location check-ins through its tight integration with geo-locating services such as Foursquare. We believe this platform could work for many other types of piggyback prototypes. Though we imagine that other platforms may be just as suitable. Facebook and Reddit are examples of platforms on which users can message each other, and thus provide an

infrastructure for piggyback prototyping. Certain limitations (such as the current $1 cost to message a non-friend on Facebook) should be considered. Each project should consider the implications of the chosen existing site. If Twitter is widely popular and accessible today, it could be different tomorrow. Piggyback prototyping would still be feasible, but a careful understanding of available social platforms is necessary.

4.5.6 What falls outside the scope of piggyback prototyping?

Not all large-scale social computing systems can be prototyped with piggyback prototyping. Three types of projects may not be well-suited to this technique: 1) those that deal with sensitive or protected data, 2) those that cannot disclose the purpose of the study to the user, and 3) those that require anonymity. For example, if the researcher has access to private data like direct messages on Twitter, then that data should not be shared with other users. Or, if the system depends on anonymity, then leveraging existing non-anonymous social networks might make it difficult to evaluate in situ. Considerations for privacy are especially important and not always straightforward. For example, we suggested that piggyback prototyping could test social algorithms such as matching algorithms. However, some algorithms might reveal information from public data that most users would not have been able to find.

4.5.7 Biases and limitations

People who publicly share broadcast messages are a self-selected group. For example, they might be more extroverted or more narcissistic. Beyond how this might impact findings in our own study on location sharing, this bias must also be considered in most piggyback prototyping systems. Second, using certain sites may not be accessible to all researchers. For our study, we used a Twitter account that was first a personal account and then evolved into a study account. As such, Twitter's automated defenses did not block it. It is possible that an account specifically created for a piggyback prototype might exhibit behaviors that would get it blocked.

4.5.8 Towards a social toolkit

In HCI, a basic building block of software UI prototyping was the development of UI toolkits that contained modular pre-defined UI components that could quickly be assembled. Could we consider the social computing systems counterpart? If we compare piggyback prototyping to the Model-View-Controller paradigm, we could use existing social data as the Model and we prototype the View and the Controller parts to varying degrees. In our example of pairing users who checked in to Twitter, we emphasized a prototype of the Controller. Others might choose to focus more on the View to prototype the visual aspect of the system. For example, Groupkit [38] provides a toolkit for videoconferencing, which facilitates the development of critical components (such as sessions) for these types of systems. Similarly, a toolkit for large social systems could be envisioned as follow-up to piggyback prototyping.

4.6 Conclusion

Piggyback prototyping is a prototyping technique for large-scale social computing systems. I described two example of using this prototyping technique: to build a plugin for more civil conversations on social media, and to build a system for pairing people checked-in to airports. The focus of the following chapter is the evaluation of the prototype for civil conversations. Based on the findings from chapter 3, I wanted to explore design alternatives to social feeds to provide more opportunities for polite conversations. Rather than building a new social media platform, recruit the participants and their social networks, and then run my evaluation, I instead employed piggyback prototyping, described in this chapter, to build the prototype. Piggyback prototyping allowed me focus on what people do on a social computing system, rather than how to attract people to the system. Next, I evaluate the prototype.

CHAPTER V

ALGORITHMIC CIVILITY PROBE DEPLOYMENT

As we saw in Chapter 3, weak ties were brittle, and people tuned out of conversations when there were too many differing opinions. Through a combination of behaviors on Facebook like hiding, tuning out, logging off, or avoiding certain conversations, people negotiated around their differences to stay connected. The challenges faced by my participants suggest ways that social platforms like Facebook could support these relationships better, through encouraging more civil behaviors.

In Chapter 4, I presented the design and implementation of a civility plugin for Facebook, following a piggyback prototyping framework. Through employing piggyback prototyping, I was able to build a prototype that lives within a user's pre-existing online social environment, and that I could deploy for an extended period of time. This allowed me to conduct a probe study that actively engaged the user with a prototype over an extended period of time.

In this Chapter, I conduct an evaluation of the civility plugin through a study methodology that is an extension of technology probe studies: an *algorithmic probe* study. Similar to a technology probe, an algorithmic probe is meant to be a rough concept that actively engages participants in a reflection around an artifact. What distinguishes algorithmic probes from others is that the probe itself, at its core, is an algorithm. End users are often consumers of the output of a social curation algorithm, and are often unaware of its inner workings, or sometimes even unaware that it is there at all [24]. Algorithmic probes engage users in reflections on decisions that should or should not be made by the algorithm.

Our rather ambitious goal was to see if there would be a change in perception of

tone on Facebook conversations after using the plugin for an extended period of time. Building a plugin of this type is complex, as we needed to leverage an infrastructure from an existing social media platform. We could have had users come into the lab to test and comment on usability aspects of the plugin. Instead, following the practice of deploying technology probes in context, we provided our users with a prototype that they used for 3 weeks in their own Facebook environment.

For this study, we employed a mixed methods approach to inform the log data of the plugin usage, and daily diaries with qualitative accounts from the participants. We found that some participants did perceive a change in civility over the course of the study. This perception change was statistically significantly correlated with increased usage of the plugin, suggesting to look more in depth at the effect of the plugin on perceptions of civility. This was also supported through the interviews. Participants who perceived a change in civility found that Facebook content was motivational and more positive during the study. They were also less concerned about missing impolite posts (through hiding) than those who did not perceive a change in civility.

These results demonstrate the usefulness of algorithmic probes as ways to inform the development of algorithms. We found that our users were able to formulate different machine learning notions (accuracy, precision, recall) as they related to their experience with the probe. We also found that encouraging pro-social behaviors, such as highlighting civil posts, is a fruitful avenue for studying social norms and deviant behavior in social media.

5.1 Research questions

The overarching research questions were:

1) Can we change people's perception of civility on social media?

Our first goal here is to explore whether there is even the potential to design social media to be more civil. We asked the study participants whether they perceived a

change in civility as they used the probe. While we did not expect our design choices for the probe to be the best way to design civil conversations in social media, seeing an indication of the potential for these designs can help refine what might be more appropriate, given that the underlying notion that improving civility in conversations can have an impact.

2) Should we design social media to encourage civility or discourage incivility?

Here we ask: What are the critical parts of our plugin that were important, or very negative on the experience of our participants? This primarily concerns itself with the interaction and visual elements of our prototype. The goal here is to determine which elements are important to consider for social platforms aiming to reduce the number of uncivil exchanges. What happened when people saw the impolite posts? Is seeing them critical to the experience? Is knowing how many were removed important? How important was the role of reflection in the experience around incivility?

3) What would be a better classifier?

We do not believe that it is necessary to have an extremely robust classifier in order to implement the features described in the previous section, however we think the politeness classifier we used could be improved. While it was outside the scope of this work to build a more robust classifier, or to improve it, through this study we were able to find a number of feature improvements, which we suggest as critical components to designing for incivility in social media. The politeness classifier mainly relied on natural language processing. We found some language, as well as behavioral characteristics, that can help pinpoint interactions deemed to be uncivil.

5.2 Technology probe

Probe studies have a variety of types in HCI. Gaver first coined the term while designing cultural probes [31]. The goal of cultural probes is to inspire emotional reactions to objects, people, or memories that carry deep personal meaning. Uncovering those "things that matter most", designers can reflect on the things that motivate people in their daily lives. Probes demand an active involvement from study participants in the research activity.

There have been a number of other types of culture probes. Some probes have the goal of inspiring experiences that drastically contrast with other known environments, such as using probes to distinguish the usage of technology in a home environment, without the technology typically used at work for productivity [28]. Urban probes place provocative technologies in urban places in order to question future technology developments in cities [72]. They emphasize the notion of rapid brainstorming as part of a probe deployment: the probe is meant to incite thought, but not be a complete design fitting into an iterative design process.

Probes can also be used to gather a specific type of information about users in a context, in order to inform the future design of systems. Cognitive probes, which incite participants to engage in reasoning activities and incite them to reflect on that cognitive process, were used in a health intervention context [57]. Others focus on logging communication data in order to better understand the type of information that might be circulated by a specific technology [28].

A technology probe encompasses many of these types of probes, as they are technology artifacts [46, 28]. Broadly speaking, a technology probe is used to evaluate novel prototypes of technological artifacts in context. With a technological probe, participants are aware of their involvement with improving the plugin. They understand that what they are using is not the complete artifact, but rather a concept.

Here we use the lens of a technology probe to frame our study approach. However, our probe differentiates itself from other instances of technology probes, in that the core component of this probe is a machine learning algorithm. In this case, we put users in control of reflecting on the accuracy of our algorithm. Our primary interest is to allow users to interact with the probe in a way that inspires them to envision a social media platform driven by our design components. The actual accuracy of the classifier is secondary provided that it lets the user understand the probe. In the traditional way of developing algorithms, user input may inform the subsequent algorithm, such as the Facebook Newsfeed algorithm, which may learn that you interact with certain friends. However, in this case the user as an active agent in the input of the system, such as providing emotional, moral, or social insights about its underpinnings, are not considered. In fact, users are often unaware of the fact that algorithms curate our online environment, or understand how an algorithm works [24].

The challenges that arise with using an off-the-shelf classifier bear some discussion at this point. A growing line of work in HCI hinges on the design of interfaces that integrate with some form of machine learning algorithm. HCI and machine learning have a long history of symbiotic work, yet we found few examples of using machine learning as a building block for a prototype. The goal of this work is not to build a high-accuracy politeness classifier; rather it is to determine whether an off-the-shelf classifier is sufficient to evaluate the interaction experience around our prototype, should a robust algorithm become available in the future. To this end, we present the notion of *algorithmic probes*, a variant of technology probes aimed at including end users in the development of machine learning algorithms.

5.3 Design and implementation of the probe

Because it would be unrealistic to study actual impacts of designing a more civil social media platform, without encountering the recruitment issues of a system trying to

obtain critical mass, we employ piggyback prototyping as a framework to develop a prototype that rests on an existing social platform. This allows us to explore the actual impact of our design choices, rather than spending time and resources trying to achieve critical mass.

The plugin has five components:

- Removing impolite posts: the entire post was hidden from the Facebook Newsfeed. Participants were not aware of what content was hidden.
- Removing impolite comments: similar to removing posts, except that participants could become aware through the conversation that certain messages were missing.
- Highlighting polite posts: changed the color of posts that were determined to be particularly polite.
- Highlighting polite comments: changed the color of comments that were determined to be particularly polite.
- Feedback about impoliteness in posts: gave feedback to the user about the impoliteness or politeness in their post.

The contrast between posts and comments was important, because they present different levels of granularity, and constitute similar building blocks to other social media platforms (instead of concentrating on Facebook-specific features).

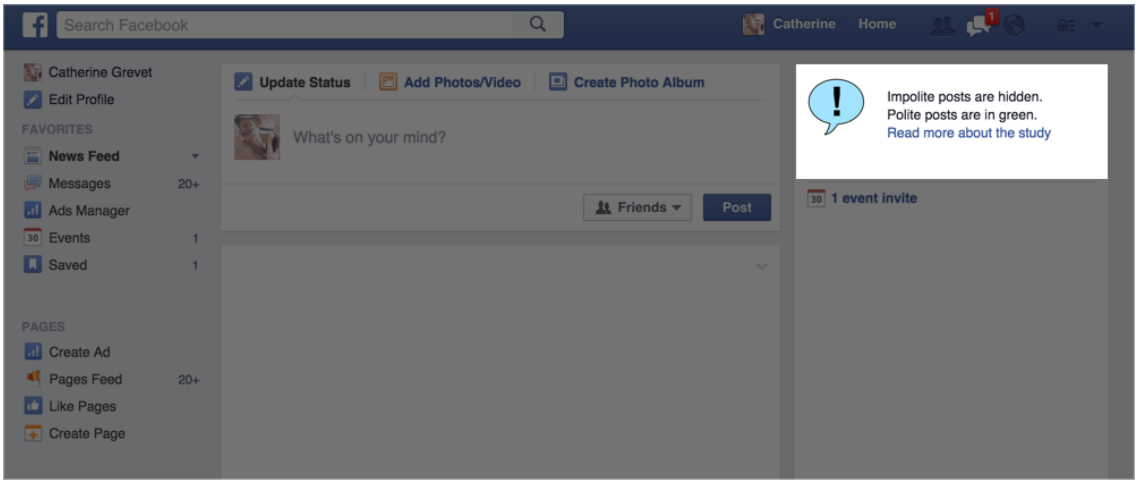

Figure 11: Screenshot of the plugin installed on Facebook.

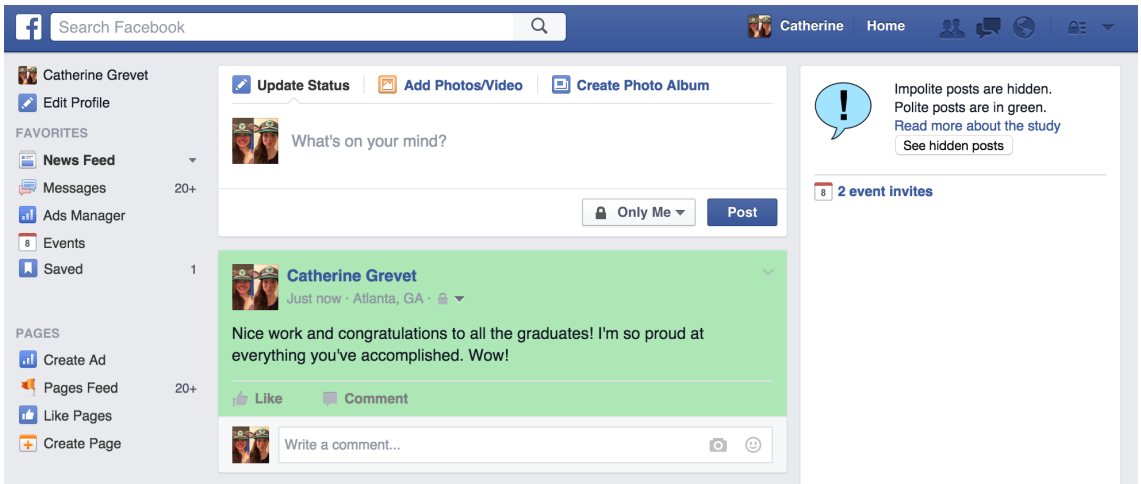

Figure 12: Screenshot of a polite post highlighted in green.

5.3.1 Building the plugin

Chapter 4 provides a more in-depth look at the technical implementation of the probe and the prototyping framework employed to build it. The front-end of the probe is a Chrome extension on top of a politeness classifier.

5.3.1.1 Classifier accuracy for Facebook

We used an off-the-shelf politeness classifier as the back-end for our prototype [16]. This classifier was trained for a specific type of sentence called requests, which are composed of a sentence and then a question [16]. We used it because it was built for social media platforms, despite the fact that not all conversations in social media are composed in a request format. The classifier features do not rest on the notion of a request structure, but rather on other aspects of the sentences. The classifier was originally trained on Wikipedia (83% accuracy) and Stack Exchange (78% accuracy) [16]. The cross-domain accuracy was lower: a classifier trained on Wikipedia was 67% accurate when classifying Stack Exchange posts; and a classifier trained on Stack Exchange was 75% accurate when classifying Wikipedia posts. This suggested that we should train a new domain before using this classifier. Since the plugin would be used on participants' Facebook Newsfeed, we trained it on Facebook data.

We trained the classifier on public Facebook posts, which are more readily available than private posts. To train the classifier, first we collected 2,557 public Facebook posts using the Facebook API for the public stream. Next, we followed the method from [16] to obtain annotated politeness scores for public Facebook posts. We sent the 2,557 collected posts to Amazon Mechanical Turk, and asked turkers to rate the politeness or impoliteness of these posts on a scale from 0 to 10. Each task had 13 Facebook posts, and each post was rated by 5 different turkers. Turkers received $0.20 per completed task, and raters were allowed to perform multiple tasks.

We trained the classifier on 1,300 of those posts, and only considered posts that

Strategy	Politeness	In top quartile	Example
1. Gratitude	0.87^{***}	$78\%^{***}$	**I** really **appreciate** that you've done them.
2. Deference	0.78^{***}	$70\%^{***}$	**Nice work** so far on your rewrite.
3. Greeting	0.43^{***}	$45\%^{***}$	**Hey**, I just tried to ...
4. Positive lexicon	0.12^{***}	$32\%^{***}$	**Wow**! / This is a **great** way to deal...
5. Negative lexicon	-0.13^{***}	$22\%^{**}$	If you're going to **accuse** me ...
6. Apologizing	0.36^{***}	$53\%^{***}$	**Sorry** to bother you ...
7. Please	0.49^{***}	$57\%^{***}$	Could you **please** say more...
8. Please start	-0.30^{*}	22%	**Please** do not remove warnings ...
9. Indirect (btw)	0.63^{***}	$58\%^{**}$	**By the way**, where did you find ...
10. Direct question	-0.27^{***}	$15\%^{***}$	**What** is your native language?
11. Direct start	-0.43^{***}	$9\%^{***}$	**So** can you retrieve it or not?
12. Counterfactual modal	0.47^{***}	$52\%^{***}$	**Could/Would** you ...
13. Indicative modal	0.09	27%	**Can/Will** you ...
14. 1st person start	0.12^{***}	$29\%^{**}$	**I** have just put the article ...
15. 1st person pl.	0.08^{*}	27%	Could **we** find a less complex name ...
16. 1st person	0.08^{***}	$28\%^{***}$	It is **my** view that ...
17. 2nd person	0.05^{***}	$30\%^{***}$	But what's the good source **you** have in mind?
18. 2nd person start	-0.30^{***}	$17\%^{**}$	**You**'ve reverted yourself ...
19. Hedges	0.14^{***}	28%	I **suggest** we start with ...
20. Factuality	-0.38^{***}	$13\%^{***}$	**In fact** you did link, ...

Figure 13: The 20 features of the politeness classifier from [16].

had 4 or more turkers rating the post as more than 5 (polite), or less than 5 (impolite). The posts left out in this case were considered "neutral", because there was not a majority consensus, and therefore were not included in the training and testing. In a 10-fold cross-validation, we obtained an accuracy of cross-validation with an average precision of 74%, and recall 73%. This significantly outperforms the baseline of 52%, representing picking the most frequent class.

We believe that a growing line of work in HCI will hinge on the design of interfaces that integrate with some form of machine learning algorithm. HCI and machine learning have a long history of symbiotic work, yet we found few examples of using machine learning as a building block for a prototype. The goal of this work is not to build the highest accuracy politeness classifier for Facebook; rather it is to determine that the off-the-shelf classifier is "good enough" in order to evaluate the interaction

experience around our prototype. By "good enough" we mean that the prototype should be sufficiently functional for participants to be able to reflect on how such an artifact would change their experience in social media.

5.4 Methods

We conducted a technology probe study using a diary and interview component with 20 participants (7 men and 14 women) over the course of 3 weeks. At the end of the study, we conducted telephone interviews. We recruited the participants through Facebook forums and Craigslist, and we encouraged snowball sampling.

5.4.1 Pre-study setup

To on-board participants, I called them and followed the following process for each of them. By email, I would send them a link to the consent form, a pre-study survey, and the code for the plugin. Over the phone, they could ask me questions about the consent form and the study protocol, then they would complete the pre-study survey, and finally I would walk them through the steps to install the plugin on Chrome. Participants were on-boarded on a rolling basis in September 2015.

5.4.2 Diary survey

At the end of each day, participants were asked to complete a daily diary. We did not require that the participants fill out the daily survey everyday, even though they did receive the email every day. The survey (see Figure 3) asked participants to rate the politeness or impoliteness of the post on a scale from 0 to 10; their relationship to the author of the post; whether they agreed with the plugin; whether they thought the message was sarcastic; and whether they would have interacted with this post on Facebook. We asked for a politeness/impoliteness score, in addition to asking whether participants agreed with how the plugin handled the post, as a way to ensure the veracity of their perception of the post, and the accuracy of the classifier.

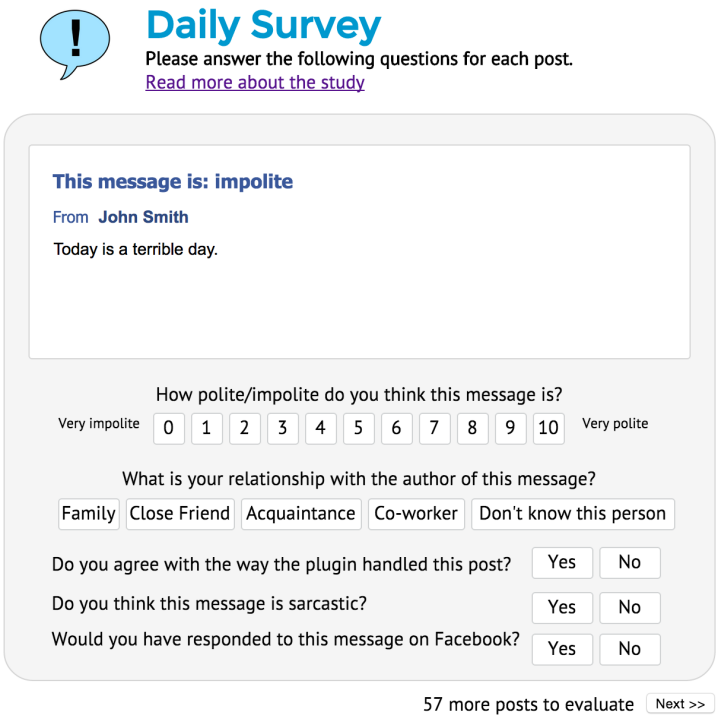

Figure 14: Screenshot of the daily diary survey.

From the usage of the plugin we obtained log data that gives us an idea of how participants used it:

- Total number of posts that were classified for each participant.
- Number of polite posts that were classified for each participant.
- Number of impolite posts that were classified for each participant.
- Number of survey responses completed by each participant, and answers to all the survey questions.

5.4.3 Post-study interview

Following the three-week study, I conducted an interview with each participant that lasted about 30 minutes. Participants were compensated by $30 for completing the final interview. The interview sought to find out whether they found anything to be different on Facebook during their study, whether their usage of Facebook had changed, how satisfied they were with what they saw on Facebook, what they thought about polite posts being highlighted in green, and impolite posts disappearing. This was a semi-structured interview, so other topics that emerged through the conversation varied depending on the participant.

Note: throughout the rest of this paper, mention of the civility "plugin" refers to the Chrome plugin + the server backend that modifies civility perception on Facebook. "The probe" refers to the plugin and the survey. These combined elements were essential elements of our technology probe deployment.

For this work we employed mixed methods. We obtained log data from the participant's usage of the probe as described above, which gave us an indication of their usage of the system. We also obtained qualitative accounts of the technology probe. The interviews were transcribed and analyzed according to themes that emerged. Results from the daily diaries complement the interviews when relevant. In addition, the results from the diary contribute to our analysis of the accuracy of the classifier. Finally, we also provide a log analysis of the posts that were collected during the study.

5.5 Perceptions of civility

While our interest in deploying this plugin as a technology probe was to see emergent themes around the notion of designing more civil social media environments, nonetheless, we were also curious to see whether the plugin would have an actual

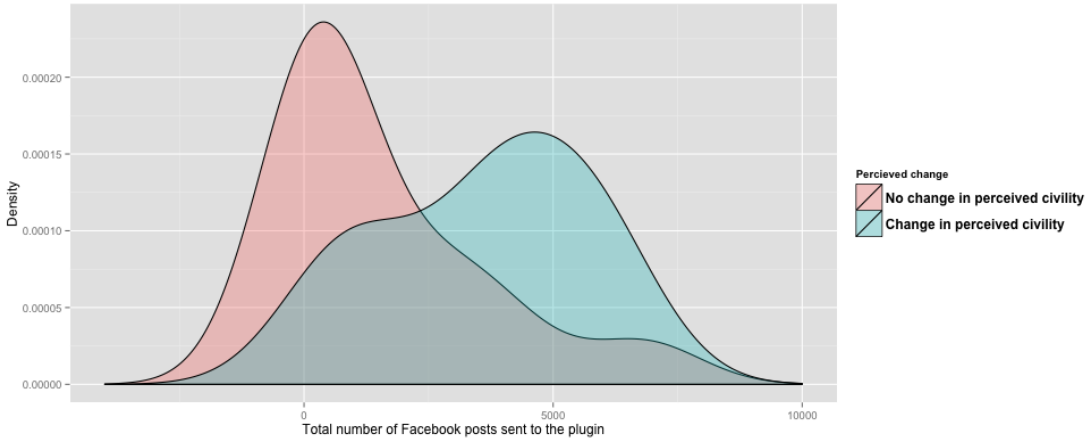

Figure 15: Density plot of the total number of posts that were sent to the classifier for those who perceived a change in civility compared to those who did not perceive a change in civility.

impact on participants' perception of civility on Facebook, in addition to fostering reflection around civility. This goal was ambitious: our design choices were rather rudimentary (in the style of a technology probe), and 3 weeks might be too short a time frame to notice any changes, and a probe deployment meant that something could go wrong at any time and impact the performance of the plugin.

From the pre-study survey, our participants reported a mix of pre-study perceptions of civility on Facebook (see Table 1). While it is uncertain how this compares to the broader population, it is clear that there was no strong bias towards participants who already perceived a lot of civility (who might already have taken steps toward a more civil experience on Facebook by unfriending people, or the such). In the post-study interview, participants were asked whether they perceived a change in civility during the study. This was asked in order to make sure that participants were providing a coherent explanation about what they had experienced, rather than gathering this information through a self-reported survey. We found that 7 out of 20 participants did perceive a positive change in civility.

The participants' perceived change in civility at the end of the study serves as the dependent variable in the rest of this Chapter. Since we saw a group that did

Table 1: Participant's self-report perception of civility on Facebook before the study.

Perception of civility	Number of participants
Very civil	0
Somewhat civil	4
Neutral	8
Somewhat uncivil	9
Very uncivil	0

perceive a change, and one that did not, we can compare these two groups to explore what might explain this difference.

5.5.1 Statistical explanation for changes in perception of civility

The differences in perceived civility are significant when looking at the total number of Facebook posts that were classified (see graphs in Figure 15). This is a proxy for the level of interaction with the plugin. Comparing the distributions of this variable shows that their difference is statistically significant. A Mann-Whitney test shows that: W = 75, p-value = 0.02136. This indicates an observable difference between those who perceived a change in civility, and those who did not, in terms of how much they used the plugin. This number was obtained by tallying the total number of incoming posts to the classifier, per participant.

From this, we can conclude that there is a correlation between the usage of the plugin and a perception of civility on Facebook. Because the assignment to usage levels was not randomized, correlation does not imply that there is a direct correlation between the plugin and a perception of civility. What aspects of the plugin contribute to this effect?

We first tried answering this through a regression model, using perceived change as the dependent variable. However, as this study contains only 20 data points when measured in this way, the statistical power of a regression model decrease for each

new variable we try to explain. A model could be determined with a follow-up large-scale study specifically focused on exploring this aspect of the probe, with a number of participants at a much larger magnitude.

Furthermore, the data obtained from the usage of the plugin created too many correlates among the results themselves. Indeed, the results obtained from the log of the plugin related to how many posts were classified, and how many were polite or impolite, and from the survey responses we knew how accurate participants found the plugin to be. However, none of these variables were particularly independent from each other. For example, the log total number of impolite posts and polite posts is correlated with the total number of posts; or the fraction of polite posts from the survey results could be correlated to the total number of polite posts. To build an accurate model, we would need to account for possible confounds.

5.5.2 Potential threats to validity

From our statistical analysis, we cannot claim which aspects of the plugin had an effect on participants' perception of civility. There are some considerations that would need to be taken into account in a future study that we suspect might have an effect on perceived civility, which are independent from the plugin.

Personal differences

There may be personality attributes or personal differences that account for a perceived change in civility. For example, one of the participants mentioned that she wouldn't use the plugin now, but she might when she was feeling more vulnerable. Perhaps accounting for mood or other characteristics of emotional state could explain some differences. Furthermore, accounting for diversity-seekers vs. challenge-averse individuals might be another variable that bears looking into [64].

In our interview results, we found that participants who might be more likely to

notice negativity were more impacted by the plugin. In particular, one participant mentioned having a tendency towards pessimism, so she liked seeing polite posts emphasized to make her see more positivity. In a future study, we could ask personality-related survey questions about participants' inclination to pessimism compared to optimism, or their openness to diversity, or whether they tend to avoid conflict.

Time effects

Perhaps events happening at the different participant enrollment times could have affected the outcome. We on-boarded participants on a rolling basis, which should account for some of this variability. However, the study was completed within a month and a half, and does not include other periods of time throughout the year. While political debates were going on at this time and not particularly different from other times of the year, there were no known major holidays during the study that could have affected Facebook usage.

Another study in this area should accommodate for this variability by controlling for timing effects, through running this study at different times during the year.

Facebook usage

Some participants mentioned using Facebook more for groups or messaging. The reasons why participants use Facebook, and other metrics about their regular interaction with the platform (maybe what time of day they usually check it, do they go to Facebook during a breaking news event, etc...) might have an effect as well. Similarly, those who use it primarily to keep in touch with family, might have a different perception from those who use it for more casual acquaintances.

Before the study, we asked participants how often they used Facebook, and from what device they accessed it. Obtaining more information about their social graph is challenging, and we did not build the plugin with the intention of collecting social

data from our participant's Facebook accounts, as this could have been prohibitive to conducting the plugin study. Other types of interactions on Facebook, such as usage of Facebook groups would be difficult to obtain without direct access to Facebook logs.

Next we examined the qualitative accounts to provide insights about the differences experienced by the participants. Through complementing the observed differences with qualitative accounts we can address questions that are not visible in the data.

Note: In the presentation of the results we will use the notation [P1, NO] or [P1, YES] to indicate the anonymous identifier of the participant, as well as whether they perceived a change in civility during the study or not. Participants with a NO are in the group of participants who did not perceive a change in civility; participants with a YES are in the group of participants who did perceive a change in civility.

5.6 Preciseness of the classifier

From the daily surveys, we can obtain a gauge of how precise our participants considered the plugin to be, given the social context of their Facebook posts. Precision is a measure of an algorithm that considers how many of the classified posts were correctly classified of those that were classified as impolite. Since we have the survey scores for the participants, we can compare their accounts of the precision of the classifier.

5.6.1 Classifier precision for polite

From the survey responses, there were 539 Facebook posts classified as polite. Of these, 470 were rated with a score above 5 from the participants, which is an agreement indicator of 87%. Most participants noticed that the plugin was good at highlighting

polite posts:

> *Yeah there's a lot of the polite comments that I agreed with. I mean it's simple compliments and things like that. You know it's worth being polite.* [P9, NO]

> *There wasn't anything that got flagged as polite that I thought wasn't polite.* [P11, NO]

5.6.2 Classifier precision for impolite

From the survey responses, there were 953 Facebook posts classified as polite. Of these, 194 were rated with a score less than 5 from the participants, which is an agreement indicator of 20%. Furthermore, 263 (27%) of them were rated neutral. And 496 (53%) were rated with a score above 5, meaning that participants considered them to be more polite than impolite or neutral. Participants reported many instances of issues with the plugin classifying posts as impolite and hiding them unnecessarily:

> *The impolite posts it's more of a neutral comment. It doesn't need have emotion or anything that shows that it's mean. And somehow it just triggers it as being impolite [...] It seems a little bit off.* [P9, NO]

In some cases, participants reported posts that were classified as impolite that they considered to be extremely polite:

> *And then if somebody says "congratulations on your new baby" those were impolite, I was like I don't understand why.* [P18, NO]

One reason that impolite posts may have been misclassified is because the classifier was originally built with the intention of classifying sentences with a sentence and request structure. In the case of this participant's example, adding an "!" to the sentence slightly improves the result, although it still registers as more impolite than

polite. The term "congratulations" is included in the positive lexicon of the classifier. It is possible that the features for the politeness classifier were better adapted to polite posts on Facebook than impolite posts. An improved civility classifier for Facebook could consider revisiting a custom feature-set for a better classifier.

5.6.3 Self-reported recall issues

Recall concerns the posts that should have been classified as polite or impolite, but were not classified as such. There were no accounts of polite posts that should have been highlighted green that were not. Neutral posts were not in the survey; thus we do not have data on participants' annotations of recall of the plugin. However, there were some concerns about certain impolite messages that were not hidden by the classifier that should have been hidden:

> *Well I think let's say, there were just a few posts that were like maybe... 7 or 8 I read that they were impolite, they should be hidden but they were posted.* [P7, YES]

Overall, through the survey results and interview feedback we found that the plugin was overly sensitive in classifying content on Facebook, and that impolite posts were especially an issue. One reason could be the fact that the threshold for impoliteness was arbitrarily set too low and was thus let too many posts through. This may be the case to some degree, but seeing that there were some polite posts that were classified as impolite is an indication that there might be something more going on. Another issue might be the fact that we need to develop more features to take into account in a context like Facebook. Bearing this out, participants were quick to note that context played a large role in their ability (or not) to assign a politeness or impoliteness score to a post.

> *You know, it's hard tell out of context. Right, so I'd have to see the rest of it to see what they were meaning. I didn't see those outright sarcastic. I'd*

have to see what came before it. Because "hey it's great weather outside", that sounds very polite. But if I don't know what the weather is like where they are, I don't know if that's sarcastic or not. [P14, NO]

In this case, the participant may have chosen a more neutral rating for the post, since they could not interpret any other context for it. Others commented on this aspect as well, claiming that it would be easier for them to judge the politeness/impoliteness of a post or comment if they saw it in the context of the rest of the conversation. These findings point perhaps to including behavioral aspects of the conversation (such as speed of back-and-forth) as additional features in the classifier.

From our statistical results, we saw a correlation between increased usage of the plugin, with a perceived change in civility on Facebook. The log analysis of the plugin was inconclusive as to what aspect of the plugin had the strongest effect. From the participant's qualitative accounts of their perception of the plugin's accuracy, they reported that polite posts were generally correctly classified, while impolite posts were often in error. This suggests that hiding impolite posts may not have been the reason behind a perceived change; and an indication that highlighting polite posts could have contributed to a positive experience with the plugin. To further understand participants' experience during the study, we provide an in-depth account of the interviews.

5.7 Experiences of participants with the probe

As mentioned earlier, 7 out of 20 of our participants perceived Facebook to be more civil after using the plugin for 3 weeks. While it is unclear from the data what exact aspect of the plugin impacted this, we can gain some insights into the impact on the plugin from the participants' personal accounts.

5.7.1 Surfacing a positive side to Facebook

From the participants' accounts, themes such as positivity and increased optimism were apparent in the ways that participants talked about the plugin, especially for those who perceived a change in civility.

5.7.1.1 Surfacing a motivational side to Facebook

This was observed in the example of two participants who noticed a change in civility while using the plugin in their comments on a motivational aspect to using Facebook that they had not encountered before:

> *It showed me that they were more motivational than I thought [laughs].* [P2, YES]

The second of these participants in particular saw an increase in posts from friends making motivational statements about going to the gym, inspiring him to go more often himself during the course of this study:

> *I was reading that... few friends were posting about how they are doing in the gym with the diet so I got a lot of information about that, I'm taking care about that, I'm getting better, it motivates me to read those posts.* [P7, YES]

Along these lines, other participants who perceived improved civility on Facebook were more attuned to noticing an uplifting aspect to Facebook, and liked the notion of highlighting positive posts in green:

> *So I liked that it would sort of highlight good news and good things. I'm looking at my Facebook now and the thing in green really standing out is that my friend's fianc just got his visa. So it's really drawing attention to this really good news. It's not hidden amongst what all the other silly things that are posted on Facebook.* [P3, YES]

5.7.1.2 Participants changed perspectives about their friends in a positive way

One participant mentions learning something new about a Facebook friend in the case where she thought that two family members were not getting along. But after seeing the post highlighted in green, she realized that there may be less tension than she imagined:

> *Ok, well for instance my mom received her Masters this past week and my sister posted a status about it and one of my aunts, my aunt by marriage, you know she commented and she was being very nice. And I think that's weird for her because she doesn't usually like my mom. But um that proved to me that she was a little bit more nicer than I thought and more happy for my mom than I thought.* [P2, YES]

It is possible that participants who were more likely to perceive a change in civility on Facebook, had common personality traits that led them to see Facebook posts more negatively when they are all lumped together in the Newsfeed. The following participant who reflects on her own tendency towards pessimism suggests how this plugin compensated for that tendency:

> *I liked that the really positive posts were highlighted, that was really nice. Like it... felt like a more positive spin on things. Like just to point out where people are being really nice to each other. [...] Well I feel like I lean towards pessimism so it was nice to... trying to think how to word this... yeah I guess just remember that some people are... I guess just remember that there are positive interactions. [...] that just was a nice reversal of I think people we have a tendency to notice the negative first um.. and so it was nice... like it visually reversed that tendency for you.* [P4, YES]

These accounts provide a concrete illustration of the benefits that participants found in using the plugin. They encountered instances during the 3 weeks of usage

that made them feel more positive, and allowed them to enjoy their social environment in ways they had not previously experienced on Facebook.

5.7.1.3 The plugin surfaces more meaningful content on Facebook

The motivational and uplifting aspects to the plugin were mentioned for the most part by the participants who had perceived a change in civility on Facebook. While those participants who did not see a change might not have experienced such drastic changes in their Facebook usage, they did tend to note that the content they were seeing on Facebook was more meaningful and interesting.

> *I did like it highlighting the polite posts. Like I found as I was just scanning through Facebook, anything that was highlighted in green I stopped and I looked at. So that certainly drew my attention to things that were more civil. I stopped and looked at them whereas I might have passed them otherwise.* [P11, NO]

> *Like I said I liked it because it showed me what to pay attention to, make sure I don't miss it. I have a lot of friends on my Facebook page, so it was nice to have a highlight. Kind of hone in on visually seeing and actually reading that particular post.* [P18, NO]

Another participant commented on seeing more posts from friends, rather than other types of content on Facebook:

> *Maybe less um... like news articles from pages that I follow and more posts from actual people maybe now that I think about it.* [P12, NO]

One participant's remark illustrates the fact that incivility on Facebook is more popular, and is often the type of behavior that is encouraged in social media, because it gets the most likes and is shared more widely. They appreciated the fact that the plugin would reverse that trend:

For one, because I got to see all the comments that I wouldn't normally see. You know because a lot of times, if people like the comments. If more people like the comments, then that's the one that you see first. But with this I see more of the meaningful comments on the posts instead of all the negative ones at the top getting all the attention. So to me it was good. [P2, YES]

Prior work has shown that incivility is more engaging than politeness, in a study of televised political debates [68]. As this participant notes, the most meaningful, positive, or motivational content is not necessarily the content that will be the most engaging (or in the case of Facebook, obtain the most likes). Since engagement is likely one consideration of the Newsfeed algorithm, it might select more controversial or "less meaningful" content than what users desire to see.

From these accounts, we see that our plugin had many opportunities to provide positive, motivational, uplifting, and interesting content to participants. They found that it made them hear more from friends, than from news streams, groups or ads, which they found to be a good thing.

5.7.1.4 Participants who perceived a change in civility saw a positive side to hidden content

Participants reconciled the fact that their Facebook Newsfeed appeared more civil as outweighing the negatives of feeling like they were missing out:

I don't know if this was the plugin sometimes when I would scroll down like things would show and then they would disappear quickly. So I don't know if those were impolite posts or it might be something with Facebook, I don't know it was weird. I did notice that. And obviously the impolite posts I never saw so yeah I probably did notice a little... slightly more positive tone. [P4, YES]

In fact, some of them described missing out as a positive and desirable aspect of the plugin:

> *I love that [it hides impolite posts] because there's a lot of things on Facebook that I wish I could hide and that I don't really want to see so I think it did great with hiding the posts.* [8903, YES]

> *You didn't really notice it while you were using it but when I went in the posts and saw the ones I missed they were kind of nice to have missed.* [P4, YES]

5.7.2 Issues with the plugin

In contrast to emphasizing polite posts, we designed impolite posts to be hidden from the Newsfeed. Participants could see them in the daily survey if they chose to complete it. Participants who did not perceive a change, described the notion of a "false sense of civility". Their perception of the fact that the plugin was hiding impolite posts was not that it contributed to a better sense of civility, but rather that it created an environment that was not representative of their friends.

> *I don't think it changed the actual civility or incivility on Facebook. I think it just gave me a false sense of the civility or incivility on Facebook.* [P11, NO]

> *What I discovered is that I'd rather see the uncivil posts in order to understand the conversation that was going on than have my sensibilities protected but not understand the conversation. I guess for me I'd rather have all the information even if it is unpleasant information.* [P11, NO]

In addition to this, participants noted other issues with hiding impolite posts. For example, it created gaps in the flow of a thread when a comment had been hidden because it was impolite.

5.7.2.1 *Missing posts from strong ties is worse than missing posts from weak ties*

Some participants commented on a nuanced aspect of civility in their conversations with close friends. They did not want to miss posts from close friends, regardless of the civility or incivility of the post.

> *Some of them, my friends, my close friends their posts were marked as impolite but they weren't.* [P14, NO]

In contrast, missing content from weaker ties was less of a concern:

> *It was an acquaintance that was responding to a posting that I responded to... they said something like "oh maybe you shouldn't have done it" or "maybe that shouldn't have happened" or something like that you know. It wasn't a long drawn out thing, you could consider it impolite, I'm not sure. [...] That was fine because like I said it was an acquaintance.* [P18, NO]

5.7.2.2 *Classification errors*

As we talked about previously, the classifier was worse at classifying impolite posts than polite posts. This might create a confound with the dislike expressed by participants of hiding impolite posts. Improving the classifier for impoliteness (and incivility more generally) is an area for future work.

Aside from improving the algorithm of the impoliteness classifier, there could also be future work focused on including user preferences into the output of the classifier. As mentioned by the following participant:

> *I do think that... I think that people should be able to choose what they want to ignore, choose whether or not something is civil. I think it someone were... if I had the ability to mark certain posts that were uncivil and like*

didn't want to see them. Then maybe the plugin could be able to recognize those particular posts maybe. But I think having someone else filter out and decide what is civil and not civil seems a little odd to me personally. [P8, NO]

This participant describes a desire to be able to determine what she considers polite or impolite. As described next, participants in the study had control over whether to have the plugin enabled (and thus moderate the experience of civility on Facebook, but there weren't more fined-grained controls to adapt the classifier on a per-user basis.

To summarize, while the fact that polite posts were highlighted in green was a positive part of the plugin, especially for participants who perceived a change in civility, the fact that the plugin hid impolite posts was generally less appealing. In fact, the removal of impolite posts created a more negative sentiment towards the plugin for those who did not perceive a change in civility. Those who perceived a change in civility were more willing to overlook the downside of missing out, in favor of the benefits of a more civil feed.

5.7.3 Sense of control

Conducting this study as a plugin gave participants the sense that they had the control to turn it on or off. Indeed, during the initial study setup, participants were walked through the steps to install the plugin, and were explained how to remove the plugin if desired. This gave participants the understanding that the plugin was an additional component to the Facebook experience that they could control. During the final interviews, some participants reflected on this quality of the experience and mentioned that there might be certain circumstances when they would want to use it, and others when they would not:

I might consider it, I guess depending on the current environment, the

political environment maybe I would consider using to block out people. [P8, NO]

Um... probably not in my current circumstances but if I were in a really rough time for some reason then I would definitely consider... like if I were feeling more fragile. I would definitely consider using it to sort of shield me a little more. [P3, YES]

Unlike the Facebook Newsfeed, which currently provides limited customizability users, this plugin - and the overall study - provided an opt-in experience around the notion of civility. This allowed participants to engage with the content they saw on Facebook with an understanding of what they were seeing, and could adapt their environment on Facebook (by choosing to uninstall the plugin for example).

5.7.4 Filtering myself vs. filtering my friends

Another aspect of the plugin was that it provided participants with feedback on their own posts. Participants didn't post very frequently, but when they did, and the classifier would consider their post impolite, they would see a popup telling them that their post is impolite.

Actually all my posts were impolite so the popup you need to make it nicer and I was just clicking ok but none of them were impolite. [P13, NO]

I thought it was funny/interesting but I didn't particularly think that whatever I was posting in that moment was uncivil. So I didn't change what I was going to post or reword it or anything based on that popup. [P11, NO]

Um I did reread it but I didn't end up changing anything. Because it was a message to my students so it was meant to be direct. But I did reread it

and be like "is this really...am I really being rude to them?" but I was like no it's not rude it's just direct so that's it. [P3, YES]

Similar to the issues that participants reported in regards to the accuracy of impolite posts, the feedback about personal posts was often questionable. This aspect of the plugin was not particularly appealing to participants.

5.7.5 Improving the plugin

Participants had a number of suggestions to improve the plugin.

5.7.5.1 Extending the plugin beyond the Facebook Newsfeed

Participants commented on applying the plugin to moderate aspects of Facebook other than the Newsfeed such as Facebook Groups or message inbox:

I would say the messages, like if you get an impolite message it could filter that. Like I think that could be really cool. Because a lot of people get you know, people bullying in their inbox and stuff. I think that might help a bit. [P6, YES]

In general, I would see the debate on my Newsfeed and then go in to the groups. And so in the past few weeks I haven't been following from the Newsfeed into the groups as often. [P3, YES]

The plugin only considered posts that were in participant's Newsfeed. This was intended to affect the central experience to Facebook usage, in order to compare the usage of the plugin between subjects. It would only require a minor change to the plugin to make it have an effect on other parts of Facebook. In the case of enabling the plugin for the Facebook message inbox, an experience around politeness and impoliteness in a one-on-one conversation scheme would require a significant change of design choices. This could be valuable future work.

5.7.5.2 Changing the color of impolite posts rather than hiding them

Many participants suggested that impolite posts should not be hidden, but rather treated similarly to polite posts by coloring them differently. Our current design is asymmetric: polite posts are rendered visually with a color, while impolite posts are hidden from the Newsfeed. The decision to hide posts was inspired by moderation strategies of online forums that remove extreme posts through hell banning, which is to hide all the posts of the offending individual. Some of our participants expressed that they would prefer to see impolite posts colored in red, or perhaps a more toned-down color.

> *Could you do it backwards. Could you like highlight in red what's really impolite? And not take anything out but just highlight it in red. Like the things highlighted in green are polite. Curse language and make that red. I wonder if that works in the opposite way. And make neutral the yellow one [laughs]. I don't know. It's like trigger warnings, if it's red then affected people wouldn't need to read it. If the whole thread has gone red, then there's probably something on there were it's really personal attacks on someone else. Or if you read it to not get too involved.* [P14, NO]

Other suggested a separate feed for impolite posts, though this would possibly result in a social media environment that would appear overwhelmingly uncivil which would not be desirable.

> *See now if it had the ability to do that, and also even... like with an anti-virus create a folder for things that it considered uncivil, or impolite, or whatever, that the person could then look through and I think that would be a great addition.* [P17, NO]

5.7.5.3 Improvements to the classifier

Participants reported that some issues with the classifier would need to be taken into account in order to improve its classification. For example, it classified acronyms as overly impolite:

> *If someone put in WTG like way to go the acronym, it marked it as impolite and I was like why? So I don't know if the plugin doesn't detect the short cuts, like LOL or whatever. A lot of them were marked as impolite which they shouldn't have been.* [P18, NO]

There were some comments about photos being impolite, and it is much more difficult to identify an offensive photo than an uncivil text comment:

> *Oh the other day one of my friends, she posted a picture of like a naked woman from the back. Like you know she wasn't from the front, she was from the back. And it was like a naked picture and she was taking a bath outside. And... she basically said to this other girl that I know, she said.. just you know basically kidding around with her. And I thought that was something that would have hidden.* [P2, YES]

5.8 Discussion

After using our probe for 3 weeks, some participants perceived more civility on Facebook than they normally see. While the statistical results don't permit us to tease apart this effect, we saw from the qualitative accounts that emphasizing polite posts was especially liked by participants who perceived a change in civility. They reported finding content on Facebook more meaningful, and discovered a motivational aspect to Facebook. Those who didn't perceive a change in civility were more concerned with missing content on Facebook due to the fact that the plugin hides impolite posts.

From their experience with the probe, participants reflected on ways to improve the accuracy of the politeness classifier and the design of the plugin. These results point to promising directions for designing more civil social media platforms.

5.8.1 Encouraging pro-social behavior

A critical aspect of these designs we wanted to explore was a contrast around discouraging impoliteness (through the probes that remove it), and encouraging civility (through visually highlighting them). There are a number of tradeoffs for both approaches, and we wanted to see what worked best for people.

Community moderation and work looking at making social platforms less harmful to participants often consider forms of discipline for undesired behavior. For example, hell banning is a common one where a troublesome person in the community is hidden from everyone else. They may not realize that no one can see what they are posting, but they eventually realize that no one is engaging with them, which should deter them, if not change their behavior.

There are components in social media that create more incentives than discipline. For example, people may collect "likes" or other engagement metrics (depending on the platform). People know what will work, or not, in their social network, in order to obtain the interest from others that they desire. In an ideal world, this might encourage people to post funny, thoughtful, enjoyable content, and to some degree probably does. However, these metrics might cater to highly provocative or uncivil content. Mutz found that people prefer televised political debates that are more uncivil, than those where polite discourse predominates [67]. Online, engagement metrics also reward content that is offensive, gossipy, violent, and having a strong shock-factor. This also encourages rumors and the spread of misinformation. Creators of such content have become experts in "click-bait", having understood what will obtain the most engagement.

In this study we saw that people experienced positive moments on Facebook when polite conversation was drawn to their attention. While it is a stretch to claim from the log data of the plugin that it was the polite posts in green that contributed to a more civil experience on Facebook, we did gain some insights about this from the qualitative accounts of participants. In particular, it emphasized meaningful content, and inspired a motivational side to Facebook that was novel to them. They found that polite posts in green drew their attention to positive content, a feature that met with appreciation.

To us, this suggests that highlighting pro-social behavior in social media is an exciting avenue for future work. This finding was more significant than the notion of hiding impolite content, showing that there may be more to gain from visually calling out pro-social behavior in social media. On Facebook or other friendship-based platforms, this could be incentive to continue with the idea of highlighting polite posts. It would be interesting to see how this work carries over to anonymous networks. For example, would we see similar conclusions for comment sections in news articles, where the focus would be on highlighting positive comments, instead of moderating the negative ones.

5.8.2 Algorithmic probes

Another notion that comes to light in this paper concerns the concept of a plugin that interacts with the Facebook Newsfeed algorithm. This plugin created an additional layer on top of the Newsfeed algorithm, that would highlight or hide posts depending on whether they were polite or impolite. It is possible that through an extended use of this plugin, there would be a feedback loop with the Newsfeed algorithm, whereby hiding impolite posts from certain people would make those people less likely to appear in the Newsfeed later in time. This supposes that it is always the same people making uncivil comments, rather than one-off commenters, and is left to future work.

Nonetheless, considering the long-term impact of such a plugin is important.

When we asked participants about their understanding of the Facebook Newsfeed algorithm, we found similar results as in [24]. Most participants were not able to explain what the Newsfeed algorithm does, none of them pointed to a potential feedback loop, and they had not thought about an interaction between the Newsfeed algorithm and the plugin, until we asked them in the interview.

In contrast, as we found in the results to the probe study, participants had many thoughts about the politeness plugin and the different components that it altered on Facebook. Using lay terms, participants were able to describe machine learning concepts, such as distinguishing between whether the plugin was better at precision or recall, and whether it seemed more accurate for politeness than impoliteness. From engaging with the plugin during the study, they were able to consider how such an algorithm would affect their social environment online. There were some suggestions, for example, to keep impolite posts visible and perhaps render them visually contrasted to polite posts. Similarly, feedback on the context of a post was important. A future algorithm might want to consider the back-and-forth of a conversation, or more about considering context to determine sarcasm.

From these thoughtful comments about the algorithm behind the plugin, we found our plugin to serve a technology probe purpose beyond inciting reflections on a novel artifact. Our probe allowed participants to actively engage with the different parameters, benefits, and downsides of an algorithm curating their social feed. Social algorithms are currently opaque to users, and we saw that our participants were not aware of how the Facebook Newsfeed works. However, they were active participants in providing an experiential account of the politeness algorithm, and improvements that could be implemented to make it better. We hope this inspires future work to consider employing algorithmic probes in order to create online social environments that are actively informed by users, rather than only altered by user input.

5.8.3 Personalization

The plugin in the study was a one-size-fits-all design, meaning that there was no personalization of the experience. Neither the plugin controls, nor the classifier, were adapted to user behavior. As the classifier was a supervised learning algorithm with no user input, all the participants in the study would have seen the same behavior of the plugin for any post that would have contained the same text. In practice, each participant had a different experience on Facebook since their friends posted different messages. Since the plugin was intended as a technology probe, it was intended to be minimal in order to let participants envision a more civil design of social media.

As we saw previously, some participants commented on a desire to personalize the output of the classifier by fine-tuning what they considered to be polite or impolite. Indeed, different people may have different tolerances or preferences for rude, crude, or profane language. For example, some might employ ample usage of cursing as a part of their interactions with their friends. They might not consider such posts as being uncivil. Future work in this area could obtain feedback from the user about these preferences.

Finally, the plugin was designed with an arbitrary threshold of politeness and impoliteness, in order to obtain three categories of posts: impolite, neutral, polite. Allowing users to set these thresholds themselves could provide an additional layer of control on the algorithm. Depending on their mood or their personal needs, they may adapt the thresholds of the plugin to skew the output of the plugin to filter more or less of the Newsfeed content.

Personalization comes with the tradeoff of a more complex interface (where should the controls be provided?) and perhaps more burden on the decision-making process of the user. While this study revealed the benefits of user control in this context, more work can focus on how to design for that sense of personalized control.

5.9 Limitations

The experimental setup for this study followed the methodology of a technology probe, every participant was given the same probe (though the differences in their social networks meant that they could have different experiences with the probe). The benefit of this setup was that all the research efforts (participant recruitment, research involvement, and analysis) could be streamlined to obtain extensive experiences with the probe from the sample population. Had I chosen a factorial experimental design, say where some participants would see a probe that would encourage civility while another group would see a probe that would discourage incivility, I would have needed to 1) have some a priori assumptions about which design variables should be distinguished, 2) forego the open-ended nature of a technology probe setup to focus on comparative results between the prototypes, and 3) recruit and on-board a larger number of participants. Given the findings from this open-ended approach, that encouraging civility led to positive experiences, a factorial experimental design approach could be a follow-up study.

As was pointed out earlier, volunteers for this study might have been biased by already being desiring a more civil experience on Facebook. Some participants reported negative reactions when they told friends about the study, so people who did not think that social media should be more civil might not have signed up for this study.

While recruiting participants on a rolling basis might account for some variability in time effects, we might see different results at a different time of year. For example, around the end of December, given many holidays and family time, people might already have very civil feeds.

Finally, I cannot control for people using Facebook from a device that does not have the plugin installed. Additionally, it is possible that they could uninstall the

plugin at some point during the day. This means that it may happen that participants do end up viewing uncivil comments. Participants should be asked before and after the study about their normal habits in accessing Facebook, such as what devices they use to access it. This should happen as an interview talking point, as it will likely highlight shortcomings of the plugin.

5.10 Contributions

In this study, 20 participants used a civility plugin for 3 weeks in their Facebook Newsfeed. This plugin was designed as an algorithmic probe, a variant of technology probes aimed to actively involve users in the reflection and developments of end-user algorithms, backed by a politeness classifier. We found that some participants did perceive a positive change civility on Facebook during the study. They expressed interest in the notion of emphasizing polite posts in green. From using this plugin, they found that posts on Facebook were more meaningful, and more motivational such as one participant going more frequently to the gym after seeing friends post about working out. In contrast, those who didn't perceive a change in civility were more concerned with missing content.

To summarize our contributions:

- We introduce *algorithmic probes* as a study methodology for actively involving users in the development of social curation algorithms. We describe this notion and present an illustrative example of its usage through a civility plugin on Facebook.

- Novel social media designs that aim to encourage civility can have a positive effect, we found that more usage of our plugin was correlated with a perception of more civility and participants were favorable of an emphasis on polite posts.

- Encouraging pro-social behaviors, in this case encouraging civility, was more

appealing to participants in this study than the opposite of hiding impolite posts. Social media platforms aiming to address deviant behavior could focus on encouraging pro-social norms rather than punishing misbehavior.

From their experience with the probe, participants reflected on ways to improve the accuracy of the politeness classifier and the design of the plugin. These results point to promising directions for designing more civil social media platforms. The implications for this work goes beyond Facebook, because civility online is a large problem for blogs and news websites. The features under study in my dissertation could also be applied to online platforms that do not rely on mutual friendship structures. For example, news websites could implement a civility reputation system, or commenters could receive feedback on the politeness of a comment they are preparing to post.

CHAPTER VI

CONCLUSION

Society benefits from sustaining a healthy deliberation process between its citizens. To this end, social media has enormous power as a platform for connecting people. Yet, as I found in my formative study, this potential for cross-cutting conversation falls short during tense political events, due to the prevalence of uncivil discourse. I found that people tune out of conversations on Facebook when political events arise, meaning that while people do have friends of opposing viewpoints, they end up not hearing their voices. My goal was to explore a design that would increase support for connections between people who do not see eye to eye.

To achieve this goal, I built a prototype, and deployed it in the context of a civility study on Facebook. The *piggyback prototype*, is a prototyping framework I envisioned for social computing projects that face issues accessing critical mass. The prototype altered participants' Facebook Newsfeeds by removing impolite posts and highlighting polite posts. This plugin was backed by a politeness classifier.

I evaluated this prototype as an *algorithmic probe* study. Participants had the plugin installed on their own Facebook page for a period of three weeks, and were actively engaged in reflecting on the functionality of the plugin. Some participants perceived an increase in civility at the end of the study, which correlated to more interaction with the plugin, than those who did not perceive a change. Participants liked seeing positive posts highlighted in green. However, impolite posts were often misclassified, and participants felt like they were missing out when those posts were removed. Overall, this work points to future directions in exploring designs to encourage pro-social behaviors.

6.1 Considerations stemming from this work

There are a number of different topics that are relevant to this thesis, but which were not directly addressed:

6.1.1 Identity vs. Anonymity

Throughout my work, I focused on studying incivility in the context of a social network based on relationship ties. On Facebook, most people know each other offline, and this means that they already have common ground and a certain frame of context for knowing how to interact with each other. We saw in this context that, despite knowing each other, there are instances in which people are rude to each other. On Facebook, these instances can be deeply personal and hurtful.

On more anonymous platforms, such as comments on news sites or other social networking sites based on pseudonyms, deviant behavior may be more widespread because of the lack of reputation between users. These platforms have often resorted to moderation in the form of banning users, hell banning, or other forms of punishment. It would be a valuable context to also study the other side of the coin: encouraging pro-social behavior. Future work could look at civility in anonymous social media, as anonymity might encourage more incivility.

6.1.2 Semantic vs. Behavioral

When we design online spaces that integrate the components developed in this dissertation, such as choosing whether to encourage or punish, there are a number of different nuances of input that could be used in the design. In particular, we should ask whether to focus on people's uncivil behaviors online, or on the content of what they post.

For example, Facebook already implements a number of behavioral patterns that shape the Newsfeed Algorithm. The content shown to users is dictated to a certain

degree (unknown publicly), by the interactions that people have on Facebook with their friends. People with whom you tend to exchange comments and photo tagging show up in your Newsfeed more often than those with whom you do not communicate as much.

The civility plugin that I studied in Chapter 5 was built on top of a politeness classifier which only considered semantics. As participants interacted with the algorithmic probe, they were able to reflect on what the classifier did well and did not do so well at classifying. They also suggested including other considerations such as tuning the algorithm based on particular people: they wouldn't want to miss any posts from close friends. More work in this area could consider incorporating a mix of semantic and behavioral features.

6.1.3 Degree of severity

Another aspect to consider is the degree of severity of the offense in online spaces. While my work mainly considers uncivil conversation, there is a broader range of deviant behavior that takes place online from minor attacks, to extreme cyberbullying. In this work, we looked in detail at incivility, a less violent form of online attacks, that might lead to larger issues societal issues such as ideological silos.

The current civility plugin may not be effective in cases of cyberbullying, where bullies may employ many methods to circumvent the politeness classifier. However, the goals of encouraging pro-social behaviors might be a relevant avenue for future work. Perhaps designing a platform in which the social norms are encouraging towards pro-social outcomes may be deterrents to extreme aggressions.

6.2 Concluding remarks

To summarize the major findings in this work:

1. People do have friends of different opinions from themselves in social media, but

they do not listen to them. We saw that during times of political contention, people would tune out of Facebook to avoid hearing from the other side, and also not share their opinion with their friends of opposing views.

2. Pro-social tools for social media can be prototyped using existing social networks. In the context of my work on breaking homophily, I repeatedly needed to build systems that require critical mass and an extensive infrastructure to run before I could even test the social science theories. I developed the piggyback prototyping framework: a plugin to make the experience on Facebook more civil. I also applied this prototype to introduce people who are different from each other on Twitter. These instances are examples of ways to implement this framework, and attest to its viability.

3. Through an algorithmic probe study, I found that a civility plugin can make participants perceive more civility on Facebook. Highlighting polite posts was particularly compelling as a design direction in this space.

6.3 Towards a Nicer Internet

A question that has persisted throughout this work combines the elements that run through all the Chapters: would it be possible to build a social layer on-top of the Internet, through a large-scale massive piggyback prototype, that would present a civil view of the Internet to those who choose to visit it? In a sense, rather than nudging people to be civil towards each other, and perhaps require some form of heavy moderating to achieve it, what if people could willingly opt in to a version of the Internet that had those capabilities in place?

For example, I could view Facebook as I normally do. Or I could view it with this Nice Internet plugin, which would leave all the interactions intact (rather than removing them like my plugin), and rewrite the content to be more polite. I could

visit CNN and read the comment section, or I could enter a "Nice CNN" where the comments would have been altered to their more civil form. Would I experience the Internet the same way that participants in my study did: more meaningful, positive, and motivational?

As described throughout this dissertation, civility refers to behaviors that act towards the well-being of a group or society. An algorithm that can understand this notion of civility is an extensive undertaking, as its goal would be to understand the content of a message in relation to a broader societal context. In this dissertation, I employed a classifier for politeness, which is one aspect of civility that can be computed through semantic and language analysis (and does not require an understanding of social context). Building a plugin for a Nicer Internet should aim towards supporting civility. This can be done by combining the learnings around designing for politeness from this dissertation with future work looking towards raising the voices of those upholding civil rights, rewarding behaviors that serve a collective good, and supporting a personalized approach to controlling one's social feed.

APPENDIX A

SURVEY FOR STUDY IN CHAPTER 3

Facebook usage questions

You may choose to log in to http://facebook.com from a desktop computer to assist you in answering these questions.

How many Facebook friends do you have? You can find the answer here: https://www.facebook.com/me/friends.*

How much do you use Facebook? * Multiple times a day — Once a day — A few times a week — A few times a month — Once a month or less

Political questions

How would you describe yourself on social political issues? * Very conservative — Conservative — Moderate — Liberal — Very Liberal — None — Rather not say

How would you describe yourself on economic political issues? * Very conservative — Conservative Moderate — Liberal — Very Liberal — None — Rather not say

Do you have friends on Facebook with differing political opinions from you about the budget cuts? *
Yes — No

Percentage of friends with different opinions

What percentage of your Facebook friends do you think have different political opinions from you about the budget cuts? (Give your best guess) * Your answer need not be exact, rather would like to know what YOU think about your network. 0 to 9% — 10% to 19% — 20% to 29% — 30% to 39% — 40% to 49% — 50% to 59% — 60% to 69% — 70% to 79% — 80% to 89% — 90% to 100%

Your Facebook usage during the Budget Cuts on March 1st, 2013

The following questions pertain to your usage of Facebook during the week following the U.S. budget cuts on March 1st. It may help to see your Facebook timeline during that period: https://www.facebook.com/me/timeline/2013/02

How much did you check Facebook during the week following the budget cuts? * Less than usual — The same as usual — More than usual

Did you join an political groups on Facebook during the week following the budget cuts? * Yes — No

Did you "like" any politician's Facebook page during the week following the budget cuts? * Yes — No

How much did you post on Facebook (status update, comment on posts, picture upload etc) during the week following the budget cuts? * Less than usual — The same as usual — More than usual

How many things did you post on your timeline (status updates, article links, photos...) during the week following the State of the Union that were relevant to the budget cuts? * 0 — 1 to 3 — 4 to 6 — 6 to 10 — More than 10

For the items you posted on your timeline that were relevant to the budget cuts... Did you set any privacy settings (such as selecting a limited audience for your post) on something you posted about the budget cuts? * Yes — No If yes, why did you set these privacy settings?

How many comments did you make on your Facebook friends' posts about the budget cuts? * You may find a list of the comments you've made here: https://www.facebook.com/me/allactivity

Was there anything you wanted to post to Facebook about the budget but didn't? * Yes — No If yes, why didn't you post these things to Facebook?

Was there anything you saw on Facebook about the budget cuts that you didn't agree with and didn't comment on? * Yes — No If yes, why didn't you comment on it?

Any other comments on your usage of Facebook during the week following the budget cuts?

Questions about specific Facebook friend

Think about one friend in particular who has a different opinion from you about the budget cuts to answer the following questions. How did you come to be friends with this person? * Childhood friend — Coworker — Family member — College friend — Other:

How many Facebook friends do you have in common with this person? (Go to this person's Facebook profile and report the number of mutual friends) * How strong is

your relationship to this person? * (barely know them) 1 2 3 4 5 (we are very close)

How much do you directly communicate with this person... * Every day: In person — Over the phone — Email — Facebook — Other
A few times a week: In person — Over the phone — Email — Facebook — Other
A few times a month: In person — Over the phone — Email — Facebook — Other
Less than once a month: In person — Over the phone — Email — Facebook — Other

If other, how else did you communicate with this person?

How much do you communicate about politics with this person in general? * Every day — A few times a week — A few times a month — Less than once a month

Did you communicate with this person during the week following the U.S. budget cuts? * Yes — No
If yes, describe your interaction with this person during the week following the budget cuts.
If no, why didn't you communicate with this person during the week following the budget cuts?

Has your relationship with this person changed after the budget cuts? * Yes — No
If yes, how has your relationship changed? Any other comments on your relationship with this person?

Demographics

Tell us more about yourself.

Gender * Male — Female — Other

Age * Younger than 18 — 18 to 30 — 31 to 40 — 41 to 50 — 51 to 60 — 61 or over

Any comments about this survey?

* means required questions

APPENDIX B

INTERVIEW QUESTIONS FOR CHAPTER 5

- How would you define the term civility?
- How important is it for conversations on Facebook to be civil?
- Did you use Facebook differently than you normally do in the last 3 weeks?
- Were there any particular events that made you use it differently?
- How often did you check Facebook?
- How satisfied were you with what you saw on Facebook in the last 3 weeks?
- How would you compare what you saw in the last 3 weeks with before you were in the study?
- Did you post anything on Facebook in the last 3 weeks? How much did you post?
- Did you see any arguments on Facebook in the last 3 weeks? Who? What happened?
- Did any relationships change during the last 3 weeks?
- Could you tell me about a time you had an argument on Facebook? Who was it with? What was it about? What happened?
- Did you talk about the plugin with anyone? Who? What did you tell them about?

- How did you feel about the plugin overall? What did you like about it? What did you dislike about it?
- Would you recommend it to a friend?
- Would you continue using it after this study?
- Would you rather this plugin be a plugin like what you used, or would you rather it be integrated into the Facebook newsfeed algorithm?
- How do you think the Facebook newsfeed algorithm works?
- Do you think this plugin interacted with the newsfeed algorithm in any way?
- Did you have any issues with the plugin?
- Does this feature do enough for you? What more should it do? How could it be improved?
- Overall, would you say this feature change the civility/incivility on Facebook in any way?

REFERENCES

[1] Adamic, L. and Glance, N., "The Political Blogosphere and the 2004 U.S. Election: Divided They Blog," in *LinkKDD*, pp. 1–16, Mar. 2005.

[2] Baumer, E., Sinclair, J., and Tomlinson, B., ""America Is Like Metamucil": Fostering Criticaland Creative Thinking about Metaphor in Political Blogs," in *CHI*, pp. 1–10, Feb. 2010.

[3] Beaudouin-Lafon, M. and Mackay, W., "Prototyping Tools and Techniques," *The Human-Computer Interaction Handbook*, pp. 1006–1031, 2002.

[4] Bernard, H. R., Killworth, P., Kronenfeld, D., and Sailer, L., "The Problem of Informant Accuracy The Validity of Retrospective Data," *Annual Review of Anthropology*, pp. 1–24, Jan. 2008.

[5] Biernacki, P. and Waldorf, D., "Snowball Sampling: Problems and Techniques of Chain Referral Sampling.," *Sociological Methods & Research*, pp. 141–163, 1981.

[6] Bishop, B., *The Big Sort: Why the Clustering of Like-Minded America is Tearing Us Apart*. Mariner Books, Nov. 2009.

[7] Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., and Fowler, J. H., "A 61-million-person experiment in social influence and political mobilization," *Nature*, vol. 489, pp. 295–298, Apr. 2013.

[8] Borah, P., "Does It Matter Where You Read the News Story? Interaction of Incivility and News Frames in the Political Blogosphere," *Communication Research*, June 2012.

[9] Burke, M. and Kraut, R., "Mind Your Ps and Qs: The Impact of Politeness andRudeness in Online Communities.," *CSCW*, 2008.

[10] Burke, M. and Kraut, R., "Using facebook after losing a job: Differential benefits of strong and weak ties," in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, (New York, NY, USA), pp. 1419–1430, ACM, 2013.

[11] Burke, M., Kraut, R., and Marlow, C., "Social Capital on Facebook: Differentiating Uses and Users," in *CHI*, pp. 1–10, Feb. 2011.

[12] Burke, M. and Kraut, R. E., "Growing closer on facebook: Changes in tie strength through social network site use," in *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, (New York, NY, USA), pp. 4187–4196, ACM, 2014.

[13] Burt, R., "Structural holes versus network closure as social capital," *Social capital: theory and research*, 2000.

[14] Conover, M. D., Ratkiewicz, J., Francisco, M., alves, B. G., Flammini, A., and Menczer, F., "Political Polarization on Twitter," in *ICWSM*, pp. 1–8, Apr. 2011.

[15] Cramer, H., Rost, M., and Holmquist, L., "Performing a check-in: emerging practices, norms and 'conflicts' in location-sharing using foursquare.," *MobileHCI*, pp. 57–66, 2011.

[16] Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., and Potts, C., "A computational approach to politeness with application to social factors," in *ACL*, pp. 1–10, May 2013.

[17] Das, S. and Kramer, A., "Self-Censorship on Facebook," *ICWSM*, 2013.

[18] DeChoudhury, M., "Tie formation on Twitter: Homophily and structure of egocentric networks.," *SocialCom.*, 2011.

[19] Dimock, Kiley, Keeter, and Doherty, "Political Polarization in the American Public," *Pew Research Center*, June 2014.

[20] Dinakar, K., Jones, B., Havasi, C., Lieberman, H., and Picard, R., "Common Sense Reasoning for Detection, Prevention, and Mitigation of Cyberbullying," *ACM Transactions on Interactive Intelligent Systems*, 2012.

[21] DiSalvo, C., *Adversarial Design*. MIT Press, 2011.

[22] Doris-Down, A., Versee, H., and Gilbert, E., "Political Blend: An Application Designed to Bring People Together Based on Political Differences," in *C&T*, Feb. 2013.

[23] Duggan, M. and Brenner, J., "The Demographics of Social Media Users — 2012," *Pew Internet and American Life Project*, Feb. 2013.

[24] Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., and Sandvig, C., ""i always assumed that i wasn't really that close to [her]": Reasoning about invisible algorithms in news feeds," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, (New York, NY, USA), pp. 153–162, ACM, 2015.

[25] Eytan Bakshy, Solomon Messing, L. A., "Exposure to ideologically diverse news and opinion on Facebook," *Science Mag*, 2015.

[26] Farrell, H., "The Consequences of the Internet for Politics," *Annual Review of Political Science*, vol. 15, pp. 35–52, June 2012.

[27] Fiorina, M. P. and Abrams, S. J., "Political Polarization in the American Public," *Annual Review of Political Science*, vol. 11, pp. 563–588, June 2008.

[28] Fitton, D., Cheverst, K., Rouncefield, M., Dix, A., and Crabtree, A., "Probing Technology with Technology Probes," *Equator Workshop on Record and Replay Technologies*, 2004.

[29] Foner, L. and Crabtree, I., "Multi-Agent Matchmaking," *Software Agents and Soft Computing: Towards Enhancing Machine intelligence, Concepts and Applications*, 1997.

[30] Garrett, R. K., "Echo chambers online?: Politically motivated selective exposure among Internet news users," *Journal of Computer-Mediated Communication*, vol. 14, pp. 265–285, Jan. 2009.

[31] Gaver, B., Dunne, T., and Pacenti, E., "Design: Cultural probes," *Interactions*, 1999.

[32] Gilbert, E., "Designing social translucence over social networks," *CHI*, 2012.

[33] Gilbert, E., "Blogs Are Echo Chambers: Blogs Are Echo Chambers," in *HICSS*, pp. 1–10, Sept. 2008.

[34] Gilbert, E. and Karahalios, K., "Predicting tie strength with social media," in *CHI*, ACM Request Permissions, Apr. 2009.

[35] Goel, S., Mason, W., and Watts, D. J., "Real and perceived attitude agreement in social networks.," *Journal of Personality and Social Psychology*, vol. 99, no. 4, pp. 611–621, 2010.

[36] Goffman, E., *The Presentation of Self in Everyday Life.* Garden City, NY: Doubleday Anchor, 1959.

[37] Granovetter, M. S., "The Strength of Weak Ties," *The American Journal of Sociology*, pp. 1360–1380, May 1973.

[38] Greenberg, S. and Roseman, M., "Groupware Toolkits for Synchronous Work," *Computer-Supported Cooperative Work: Trends in Software 7*, 1999.

[39] Grevet, C. and Gilbert, E., "Piggyback Prototyping: Using Existing, Large-Scale Social Computing Systems to Prototype New Ones," *CHI*, 2015.

[40] Grevet, C., Terveen, L., and Gilbert, E., "Managing Political Differences on Facebook," *CSCW*, 2014.

[41] Grudin, J., "Groupware and social dynamics: eight challenges for developers," *Communications of the ACM*, 1994.

[42] Guy, I., e. a., "Same places, same things, same people?: mining user similarity on social media.," *CSCW*, pp. 41–50, 2010.

[43] Hancock, J., Toma, C., and Fenner, K., "I Know Something You Don?t: The Use of Asymmetric Personal Information for Interpersonal Advantage.," *CSCW*, 2008.

[44] Hemphill, L., Otterbacher, J., and Shapiro, M., "What's Congress Doing on Twitter?," *CSCW*, pp. 1–10, Feb. 2013.

[45] Hetherington and Weiler, *Authoritarianism and Polarization*. 2009.

[46] Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B. B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, S., Evans, H., Hansen, H., Roussel, N., and Eiderbäck, B., "Technology probes: Inspiring design for and with families," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, (New York, NY, USA), pp. 17–24, ACM, 2003.

[47] Iachello, G., "Prototyping and sampling experience to evaluate ubiquitous computing privacy in the real world.," *CHI*, pp. 1009–1018, 2006.

[48] Irani, L. and Silberman, S., "Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk," *CHI*, 2013.

[49] Jisun, A., Quercia, D., and Crowcroft, J., "Fragmented Social Media: A Look into Selective Exposure to Political News," in *WWW*, pp. 1–2, Apr. 2013.

[50] Kelly, N., "Can Democrats Get Republicans to Move on Scalia's Replacement?." http://www.theatlantic.com/politics/archive/2016/02/senate-democrats-republicans-scalia-supreme-court/471084/, 2016.

[51] King, ML, J., "Letter from Birmingham Jail," 1963.

[52] Kramer, A. D., "The spread of emotion via facebook," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, (New York, NY, USA), pp. 767–770, ACM, 2012.

[53] Kriplean, T., Morgan, J., Freelon, D., Borning, A., and Bennett, L., "Supporting Reflective Public Thought with ConsiderIt," pp. 1–10, Feb. 2012.

[54] LaBarre, S., "Why We're Shutting Off Our Comments." http://www.popsci.com/science/article/2013-09/why-were-shutting-our-comments, 2013.

[55] Lapides, P., Chokshi, A., Carpendale, S., and Greenberg, S., "News feed: What's in it for me?," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, (New York, NY, USA), pp. 163–172, ACM, 2015.

[56] Lindqvist, J., e. a., "I'm the mayor of my house: examining why people use foursquare - a social-driven location sharing application.," *CHI*, pp. 2409–2418, 2011.

[57] Mamykina, L., Mynatt, E. D., and Kaufman, D. R., "Investigating health management practices of individuals with diabetes," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, (New York, NY, USA), pp. 927–936, ACM, 2006.

[58] Matias, J. N., Johnson, A., Boesel, W. E., Keegan, B., Friedman, J., and DeTar, C., "Reporting, Reviewing, and Responding to Harassment on Twitter," 2015.

[59] McDonald, D., Gokhman, S., and Zachry, M., "Building for social translucence: a domain analysis and prototype system," *CSCW*, 2012.

[60] McPherson, M., Smith-Lovin, L., and Cook, J. M., "Birds of a feather: Homophily in social networks," *Annual review of sociology*, pp. 415–444, 2001.

[61] Milgram, S., *The Familiar Stranger: An Aspect of Urban Anonymity.* 1972.

[62] Mitchell, Gottfried, Kiley, and Matsa, "Political Polarization and Media Habits," *Pew Research Center*, Oct. 2014.

[63] Mouw, T. and Sobel, M. E., "Culture Wars and Opinion Polarization: The Case of Abortion," *American Journal of Sociology*, pp. 1–31, Apr. 2001.

[64] Munson, S. and Resnick, P., "Presenting Diverse Political Opinions: How and How Much," in *CHI*, pp. 1–10, Apr. 2010.

[65] Munson, S. and Resnick, P., "The Prevalence of Political Discourse in Non-Political Blogs," in *ICWSM*, pp. 1–8, July 2011.

[66] Mutz, D., "The Consequences of Cross-Cutting Networks for Political Participation," *American Journal of Political Science*, vol. 46, pp. 838–855, Oct. 2002.

[67] Mutz, D. and Martin, P., "Facilitating Communication across Lines of Political Difference: The Role of Mass Media," *American Political Science Review*, 2001.

[68] Mutz, D. and Reeves, B., "The New Videomalaise: Effects of Televised Incivility on Political Trust," *American Political Science Review*, pp. 1–16, Feb. 2005.

[69] Nichols, J. and Kang, J.-H., "Asking questions of targeted strangers on social networks.," *CSCW*, 2012.

[70] Papacharissi, Z., "Democracy online: civility, politeness, and the democratic potential of online political discussion groups," *New Media & Society*, 2004.

[71] Pater, "Something Awful," *CHI*, 2014.

[72] Paulos, E. and Jenkins, T., "Urban probes: Encountering our emerging urban atmospheres," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, (New York, NY, USA), pp. 341–350, ACM, 2005.

[73] Putnam, R., "Bowling alone: the collapse and revival of American community," *Simon & Schuster, New York*, 2000.

[74] Putnam, R. D., "E Pluribus Unum: Diversity and Community in the Twenty-first Century The 2006 Johan Skytte Prize Lecture," *Nordic Political Science Association*, 2007.

[75] Rader, E. and Gray, R., "Understanding user beliefs about algorithmic curation in the facebook news feed," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, (New York, NY, USA), pp. 173–182, ACM, 2015.

[76] Rainie, L. and Smith, A., "Social networking sites and politics," *Pew Internet and American Life Project*, Mar. 2012.

[77] Rainie, L., Smith, A., Scholzman, K., Brady, H., and Verba, S., "Social Media and Political Engagement," *Pew Internet and American Life Project*, pp. 1–13, Oct. 2012.

[78] Rettig, M., "Prototyping for tiny fingers," *Communications of the ACM*, 1994.

[79] Sleeper, M., Balebako, R., Das, S., McConahy, A. L., Wiese, J., and Cranor, L. F., "The Post that Wasn't: Exploring Self-Censorship on Facebook," *CSCW*, pp. 1–10, Feb. 2013.

[80] Smith, C., "Conduct Control on Usenet," *JCMC*, 1997.

[81] Stutzman, F., Gross, R., and Acquisti, A., "Silent Listeners: The Evolution of Privacy and Disclosure on Facebook," *Journal of Privacy and Confidentiality*, vol. 4, pp. 7–41, Mar. 2012.

[82] Sukumaran, A., Vezich, S., McHugh, M., and Nass, C., "Normative Influences on Thoughtful Online Participation," in *CHI*, pp. 1–10, May 2011.

[83] Sunstein, C., "The Law of Group Polarization," *The Journal of Political Philosophy*, vol. 10, pp. 175–195, Apr. 2002.

[84] Sunstein, C., "Republic.com 2.0.," *Princeton University Press*, 2009.

[85] Tang, K.P., e. a., "Rethinking location sharing: exploring the implications of social-driven vs. purpose-driven location sharing.," *Ubicomp*, pp. 85–94, 2010.

[86] Terveen, L. and McDonald, D., "Social matching: A framework and research agenda. ," *ACM Trans. Comput.-Hum. Interact.*, 2005.

[87] Wojcieszak, M. E. and Mutz, D. C., "Online Groups and Political Discourse: Do Online Discussion Spaces Facilitate Exposure to Political Disagreement?," *Journal of Communication*, vol. 59, pp. 40–56, Mar. 2009.

[88] Zhao, C., Hinds, P., and Gao, G., "How and To Whom People Share: The Role of Culture in Self-Disclosure in Online Communities," in *CSCW*, pp. 1–10, Nov. 2012.